Psychological Bulletin

CONTENTS

Published Bimonthly by the American Psychological Association

Vol. 53, No. 3 May, 1956

The Psychological Bulletin contains evaluative reviews of research literature and articles on research methodology in psychology. This JOURNAL *does not publish reports of original research or original theoretical articles.*

Manuscripts should be sent to Wayne Dennis, Department of Psychology, Brooklyn College, Brooklyn 10, New York. Manuscripts should be sent to this address even during the editor's absence on leave. Manuscripts will be acknowledged at this address and the carbon copies forwarded to the editor. However, beginning on August 1, 1955 and until August 1, 1956 *correspondence* should be addressed to the editor at the Department of Psychology, American University, Beirut, Lebanon.

Preparation of articles for publication. Authors are strongly advised to follow the general directions given in the "Publication Manual of the American Psychological Association" (*Psychological Bulletin*, 1952, 49 [No. 4, Part 2], 389–449). Special attention should be given to the section on the preparation of the references (pp. 432–440), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted *in duplicate.* Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

Reprints. Fifty free offprints are given to contributors of articles and notes. Authors of early publication articles receive no gratis offprints.

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 15th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Entered as second class mail matter at the post office at Washington, D.C., under the act of March 3, 1879. Additional entry at the post office at Menasha, Wisconsin. Acceptance for mailing at special rate of postage under the provisions of Sec. 34.40 Par. (D) provided for in Section 5.34, act of February 28, 1925, authorized August 6, 1947. Printed in U.S.A.

VOL. 53, NO. 3 MAY, 1956

Psychological Bulletin

INTELLIGENCE AND FAMILY SIZE

ANNE ANASTASI
Fordham University

The past twenty-five years have witnessed a growing interest on the part of psychologists, geneticists, and demographers in the relationship between intellectual level and family size. As a result of the negative correlations commonly found between intelligence test scores of children and number of siblings, several writers have predicted a drop in the intellectual level of the population. Data which have recently become available from a variety of sources, however, raise serious doubts regarding such a conclusion. Careful analyses of the problem, moreover, have revealed that it is far more complex —both methodologically and theoretically—than was originally supposed.

THE GENERAL QUESTION OF DIFFERENTIAL FERTILITY

In current English demographic usage, the term "fecundity" signifies capacity to produce living offspring, whereas "fertility" refers to actual performance (**60**). Some investigators have differentiated further among number of pregnancies, number of births (including stillbirths), number of children born alive, and number of children alive at the time of the survey (**71**). For practical reasons, however, most studies have been concerned only with the last-named category. To be precise, one should also identify foster children and stepchildren within each family, but not all studies have done so. The number of such cases, of course, is relatively small and would not materially affect the conclusions.

Mention should also be made of the frequently reported "crude birth rate," which is simply the number of births during a year, per thousand population. Such a figure reflects not only mean size of family but also other demographic characteristics, such as the age distribution, sex ratio, and marriage habits of the population. Adjusted birth rates are sometimes computed, in which one or more of these characteristics are controlled. On the other hand, when average size of sibship is considered, as in most studies on fertility and intelligence, no account is taken of childless families, proportion of unmarried persons, and differential mortality rates. It is thus apparent that fertility statistics can be variously expressed and that their specific implications need to be carefully scrutinized.

Most demographers agree that, although capacity for procreation may be influenced by genetic factors and by such environmental conditions as nutrition and climate, variations in fertility among existing human populations arise chiefly from social and psychological factors (**60, 96**). Similar explanations have been offered for progressive changes in fertility within given populations. In the United

States and in most European countries, the birth rate declined steadily from the early or late nineteenth century to the 1930's, reaching its lowest point during or immediately following the economic depression. This decline has usually been attributed to socioeconomic factors, such as urban migration, the rising standard of living, and the increasing expense of rearing children (**96**). Following World War II, the trend was reversed, the crude birth rate for these countries rising markedly and reaching a peak in the late 1940's. Much of this rise is believed to be due to the abnormally large number of marriages during and immediately after World War II, as well as to births which had been postponed during the depression or early part of the war, or which were advanced from later years because of the economic prosperity (**96**). Some demographers, however, see evidence of a genuine increase in size of completed families (**96**).

Differential fertility in various subdivisions of a population was recorded as early as the seventeenth century in Europe (**96**). In a survey of British demographic writings appearing between 1660 and 1760, for example, Kuczynski (**52**) found references to the greater fertility of rural as compared with urban dwellers, and of the poor as compared with the wealthy. The data upon which these early opinions were based are, of course, meager and difficult to interpret. In more recent times, however, a mass of data has accumulated which shows an inverse relationship between family size and such variables as income, occupational level, and amount of education (**47**, **61**, **84**, **96**). The previously mentioned decline in birth rate which began in the nineteenth century was more rapid in the upper than in the lower socioeconomic and educational classes, thus either producing or augmenting such fertility differentials.

Certain other findings regarding fertility differentials are of special interest. One of the most persistent differentials, occurring even when other subgroup differences in family size are absent or reversed, is that between urban and rural groups. This difference usually increases when the comparison is between agricultural and nonagricultural occupations. The largest families are thus found among those rural residents who are engaged in agriculture (**96**). Another suggestive finding, reported in an American survey, is that family size tended to be more closely related to the educational level of the wife than to that of the husband (**5**). Finally, it is noteworthy that a comparison of native whites and Negroes in the United States reveals that the fertility rates of both groups are affected in the same way by such factors as socioeconomic status, education, and urban-rural residence (**53**).

All of the above socioeconomic and educational differences in fertility have been interpreted as indirect evidence for a relationship between fertility and intelligence. Insofar as occupational and educational level are positively correlated with intelligence test scores, and urban scores average higher than rural, the inverse relation of these variables with family size suggests a similar inverse relation between intelligence and family size. Even prior to the direct study of the latter relationship, therefore, its implications had been considered by geneticists and demographers.

There are, however, a number of noteworthy exceptions to the relationships cited above. For example, several local studies in American, Canadian, French, German, Nor-

wegian, and Swedish cities have shown that the negative correlation between income and family size prevails only up to a certain income level (**96**). Moreover, within the higher income groups, there is sometimes a tendency for family size to be positively correlated with income. Similarly, exceptions can be found to the hypothesized relation between fertility and occupational level (**30**).

Special mention should be made of the Indianapolis survey, conducted in 1941–42 (**48, 49, 50, 51, 106, 107**). This study began with a short questionnaire administered to approximately 41,500 native, white couples living in Indianapolis, in which the wife was under 45 years of age and neither spouse had previous marriages. The major data were gathered in an intensive interview survey of about 1,500 women selected from the larger sample because they met certain additional specifications. The investigation was designed to test a number of hypotheses regarding social and psychological factors affecting fertility. Many of the findings are difficult to interpret, however, because of the complex interrelationships and interactions of the variables concerned (cf. **48, 51**). Nevertheless, one observation is of special interest in the present connection. Although the usual inverse relationship between socioeconomic indices and fertility was found over a large part of the range, this relationship changed from negative to positive within the highest income levels. Such a result held in both the total sample and the subsample chosen for intensive study. It was likewise found to be true in another subsample including only wives aged 40 to 44 years, who were more likely to have completed families at the time of the survey (**50**).

One of the hypotheses proposed to account for such results is based upon the relation of upward social mobility to family size (**104**). There is some direct evidence in support of this hypothesis. Baltzell (**6**), for example, found larger families among residents of Philadelphia listed in both the 1940 *Who's Who* and the *Social Register* than among those listed only in *Who's Who*. He argued that the latter had more often achieved their socially prominent position through their own efforts rather than having occupied it throughout life. They had thus shown more upward mobility than the *Social Register* group. Other analyses conducted within the same study, involving persons with a private school education and those whose families had been listed in *Who's Who* since 1900, likewise indicated greater fertility among the socially less mobile groups. Attempts to test the hypothesis among intermediate socioeconomic groups have yielded inconsistent results. At such levels, it is difficult to find groups in which the drive for upward social mobility can be assumed to be negligible. For example, persons whose occupational levels are higher than those of their fathers or who have advanced in occupational level within their own vocational history have been compared with those who failed to show such changes (**7**). Even the "control" group which has manifested no occupational changes of this type, however, may have manifested upward mobility in other ways.

It is apparent from various sources that, even prior to the postwar "baby boom," generalizations regarding the inverse relationship of fertility to socioeconomic factors needed qualification. Following World War II, fertility increased relatively more among those groups in which the prewar level had been low, i.e., among urban residents and persons in higher

educational, occupational, and income levels. Thus the increasing birth rate has had the effect of narrowing fertility differentials within the population. These changes are vividly illustrated by some of the data gathered in recent sample surveys by the United States Bureau of the Census (**98, 99**). For example, when married women between the ages of 15 and 44 were considered in the 1952 survey, no relationship was found between income level and total number of children, except for a proportionately large number of children in the lowest income group (under $1,000). When the same women were classified according to husband's occupation, some reversals in the usual fertility differentials were observed. Thus, professional, semiprofessional, and managerial groups had proportionately more children than did certain lower occupational groups, such as clerical and sales workers.[1] Fertility differentials among educational groups have also decreased (**98**). The remaining differences between educational classes are still large, however, as are those between urban and rural groups.

A similar narrowing of fertility differentials has been reported for a number of European countries (cf. **16, 96**). For example, there is evidence of an actual increase in family size in the upper social levels of Sweden and Great Britain. Moreover, the decrease in fertility differentials holds for comparisons between countries as well. Thus it is largely in the low-fertility countries that the birth rate has increased.

Mention may also be made of conditions prevailing within high-fertility areas. These are identified as regions having a crude birth rate in the vicinity of 40 per thousand, or higher (**96**). Such birth rates have been reported among peoples of Africa, Asia, and Latin-America; among non-European peoples of Oceania; and in certain parts of Europe. For most of these groups, demographic data are very inadequate. In general, the birth rate in these areas has shown no upward or downward trend, unlike the birth rate in low-fertility areas. Moreover, the usual fertility differentials have not been observed in high-fertility areas. In Brazil, for example, the urban-rural differential is the only one which follows the familiar pattern. Reported socioeconomic differentials in that country are slight and not in the expected direction. Thus within each type of work, the fertility of employers and own-account workers is generally higher than that of employees and unpaid family workers (**96**). The data from high-fertility areas, however meager, provide another major exception to the negative correlation between socioeconomic level and family size.

Studies of Intelligence Test Performance in Relation to Family Size

While much has been inferred about the relationship between fertility and intelligence from studies on socioeconomic factors, a direct approach to this problem has also been utilized. Current concern with this question on the part of demographers is illustrated by the inclusion of a session on "methods of research on relations between intelligence and fertility" in the World Population Con-

[1] The fact that significant differences in family size were obtained among occupational groups, while no such differences were found among income levels within the corresponding range, reflects the low correlation between income and occupational level. The range of income within each of the major occupational categories is very wide, and the income differential between such occupational groups has been declining for many years.

ference held under United Nations auspices in Rome in 1954 (**97**). Similarly, a special working group was called together by Unesco to discuss the implications of fertility differentials for the intellectual development of the population. This group met in Paris early in 1954 in order to draw up a statement which was contributed by Unesco to the World Population Conference (**17**).

Investigations on the relationship between family size and intelligence test score have been conducted in several countries, although the liveliest interest in this problem has been manifested in Great Britain. Among such studies, the outstanding examples are the Scottish surveys (**81**, **82**, **83**). In each of these surveys, an effort was made to test a complete age sample. In 1932 and again in 1947, a 45-minute group intelligence test including verbal and pictorial material was administered to all 11-year-old children in Scotland. The samples actually tested consisted of 87,498 and 70,805 children in the first and second surveys, respectively. These samples are described as complete except for the children whose sensory or motor handicaps precluded the valid administration of this test, those school children who were absent on the day of testing, and a few children attending certain private schools for whom the necessary background data could not be obtained. In terms of the total population of 11-year-old Scottish children, the two samples afforded approximately the same coverage. Since the estimated number of 11-year-old children in Scotland was 100,300 in 1932 and 80,300 in 1947, the percentages covered in the two samples were 87 and 88, respectively.

Surveys with individual tests were likewise conducted on more narrowly defined samples. All children born in Scotland on the first day of February, May, August, or November, 1926 were given the 1916 Stanford-Binet and a series of eight performance tests (**63**). This group comprised 874 cases, tested over a three-year period at ages ranging from 8–11 to 11–9. In connection with the 1947 group test survey, 1,215 children were also tested with Form L of the 1937 Stanford-Binet, as modified for Scottish children. This group, known as the "six-day sample," included children born in Scotland on the first day of the even months in 1936. The coverage of the individual test surveys is more nearly complete than that of the group test surveys, special efforts having been made to reach every child who met the birthday specifications. Moreover, the 1,215 children constituting the "six-day sample" of the 1947 survey were studied more intensively in terms of background characteristics. The same children are currently being followed up to determine the effect of family size and other factors upon their subsequent development (**66**).

In the earlier Scottish surveys, no data on family size were obtained, although the results of these surveys were employed in the follow-up studies to be discussed in a later section of this paper. In the 1947 survey, the group test[2] yielded a correlation of −.28 with size of sibship, and the Stanford-Binet a correlation of −.32. Comparisons of mean scores likewise showed a consistent and significant drop with increase in family size. Thus the mean Stanford-Binet IQ dropped steadily from 113 for "only" children to 91 for children with five sibs (**82**). On the group test, the

[2] Only the verbal part of the group test was included in this analysis. The results obtained with the pictorial part proved to be too highly skewed because of low test ceiling; hence they have not been analyzed further.

mean decrease in score per unit-increase in family size was .13σ (**83**).

The tendency for children from larger families to score lower persisted within occupational classes (**83**), a finding which has also been reported by other investigators (**14, 21**). All such studies, however, have employed fairly broad occupational categories. Thus it cannot be assumed that socioeconomic differences have been ruled out when comparisons are made within single occupational levels. It is also interesting to note that, in the highest occupational levels, the negative correlation between intelligence and family size may disappear or be replaced by a positive correlation. Cattell (**21**) and Moshinsky (**68**) report that the highest negative correlations are found in occupations of intermediate social status. In this connection, reference may again be made to the factor of upward social mobility, discussed in the preceding section. It is at the intermediate social levels that such a factor would be most likely to operate.

In the 1947 Scottish survey, group test scores were available for 974 twins (**83**). Analysis of these scores corroborated the usual finding that twins score lower than singletons, the difference in this group corresponding to approximately 5 IQ points. It was further demonstrated that such a difference could not be attributed to family size or to socioeconomic level. There was no tendency for the twins to come from larger families, when number of births rather than number of children was considered. Such socioeconomic indices as relative size of home and paternal occupation indicated no inferiority of the twin sample. In fact, there was an excess of twins in the professional and employer groups, and a deficit in the unskilled labor group. The implications of these findings will be considered below, after other pertinent data have been examined.

Another extensive survey is that conducted among 6- to 12-year-old French school children in 1943–44 (**36, 38**). This survey included 95,237 boys and girls, approximately a 2 per cent sample of the elementary school population of France. Care was taken to obtain a representative distribution of cases over the twenty regions of France. All subjects were given René Gille's "test Mosaïque," a specially constructed pictorial group test of intelligence.[3] When analyzed in respect to family size, the mean test scores exhibited a consistent drop with increasing size of sibship. This differential was apparent within each year group, the difference in mean score between a one-child family and a family of eight or more being equivalent to 1 or 2 years of mental age. Within occupational groups, the negative relationship between intelligence and family size was clearly apparent among farmers, manual laborers, and clerical workers; it was barely discernible in the managerial group and negligible in the professional class.

A special study was concerned with the length of intersibling interval (**36**, Ch. 3; **90**). Within the total sample covered by the French survey, there were 1,244 two-sibling families in which both siblings had been tested. These were separated into "long interval" and "short internal" sibships, the latter being defined as those falling at or below

[3] This test is described and completely reproduced in the first report of the French survey (**38**, pp. 75–96). Data on reliability and validity are included in the second report (**36**, pp. 21–45). It was called the "mosaic" test because it includes a wide variety of items.

the median interval. In the urban group the median interval was 24 months, in the rural 23 months. On the intelligence test, the children with long intersib intervals obtained significantly higher means, these differences persisting within each of the five occupational categories into which the sample was subdivided. With long intersib intervals, the scores approximated those of only children; with short intersib intervals, they approximated the scores obtained by three-child sibships.

Test scores were also available for 750 twins (**36**, Ch. 3; **90**). In this survey, family size averaged higher for twins than for the total sample, even when the twins were counted as a single birth. Analysis of paternal occupation showed a greater proportion of farmers in the twin sample than in the total sample. With regard to all other occupational classes, however, the twin sample was superior to the total sample. Thus a higher percentage of twins had fathers in the managerial and professional classes, and a lower percentage in the clerical and manual labor categories than was the case in the total sample. On the whole, then, the findings pertaining to occupation are similar to those of the Scottish survey, in which paternal occupation was somewhat superior for twins than for singletons.

The test scores of the twins were again inferior to those of the total sample. This difference was found at each age, and remained when family size and occupational level were controlled. In the case of families of four or more children, however, the twin inferiority was negligible and inconsistent, being most pronounced in the case of two- and three-sib families. Moreover, among the two-sib families, the twin scores were very similar to those of short-interval sibs, as defined in the earlier part of the study.

Other similar surveys have been conducted on a less extensive scale in several countries. In England, there have been a number of attempts to test complete age samples of children within restricted areas. Fraser Roberts and his co-workers (**77**, **79**) administered the Otis test to all children born during a four-year period and living within the city of Bath; a subsample was also individually tested with the Stanford-Binet. The Cattell Culture-Free Test was used by Cattell (**18**, **19**) with all 10-year-old children in Leicester, a typical industrial city, and Devonshire, an "unspoilt rural area." Sutherland and Thomson (**89**) applied a specially developed group intelligence test to all 11-year-old school children in the Isle of Wight. Burt (**14**) reports the results obtained with the elementary school population of one of the London boroughs in the course of his standardization of the Binet scale. The Otis test scores of 393 10-year-olds in Sheffield were analyzed by Bradford (**12**). All of these studies confirm the findings of the more extensive surveys regarding the decline in mean intelligence test score with increasing size of sibship. When correlations were computed between test score and family size, the coefficients were uniformly negative and clustered in the .20's and low .30's.

A similar inverse relation between intelligence and family size is reported by Papavassiliou (**72**), who gave a Greek adaptation of the Stanford-Binet to 349 children in Athens. Heinen (**37**) found a tendency for family size to be inversely related to school grades and other indices of academic ability in several large samples of German school children. In another survey of German school

mean decrease in score per unit-increase in family size was .13σ (**83**).

The tendency for children from larger families to score lower persisted within occupational classes (**83**), a finding which has also been reported by other investigators (**14**, **21**). All such studies, however, have employed fairly broad occupational categories. Thus it cannot be assumed that socioeconomic differences have been ruled out when comparisons are made within single occupational levels. It is also interesting to note that, in the highest occupational levels, the negative correlation between intelligence and family size may disappear or be replaced by a positive correlation. Cattell (**21**) and Moshinsky (**68**) report that the highest negative correlations are found in occupations of intermediate social status. In this connection, reference may again be made to the factor of upward social mobility, discussed in the preceding section. It is at the intermediate social levels that such a factor would be most likely to operate.

In the 1947 Scottish survey, group test scores were available for 974 twins (**83**). Analysis of these scores corroborated the usual finding that twins score lower than singletons, the difference in this group corresponding to approximately 5 IQ points. It was further demonstrated that such a difference could not be attributed to family size or to socioeconomic level. There was no tendency for the twins to come from larger families, when number of births rather than number of children was considered. Such socioeconomic indices as relative size of home and paternal occupation indicated no inferiority of the twin sample. In fact, there was an excess of twins in the professional and employer groups, and a deficit in the unskilled labor group. The implications of these findings will be considered below, after other pertinent data have been examined.

Another extensive survey is that conducted among 6- to 12-year-old French school children in 1943–44 (**36**, **38**). This survey included 95,237 boys and girls, approximately a 2 per cent sample of the elementary school population of France. Care was taken to obtain a representative distribution of cases over the twenty regions of France. All subjects were given René Gille's "test Mosaïque," a specially constructed pictorial group test of intelligence.[3] When analyzed in respect to family size, the mean test scores exhibited a consistent drop with increasing size of sibship. This differential was apparent within each year group, the difference in mean score between a one-child family and a family of eight or more being equivalent to 1 or 2 years of mental age. Within occupational groups, the negative relationship between intelligence and family size was clearly apparent among farmers, manual laborers, and clerical workers; it was barely discernible in the managerial group and negligible in the professional class.

A special study was concerned with the length of intersibling interval (**36**, Ch. 3; **90**). Within the total sample covered by the French survey, there were 1,244 two-sibling families in which both siblings had been tested. These were separated into "long interval" and "short internal" sibships, the latter being defined as those falling at or below

[3] This test is described and completely reproduced in the first report of the French survey (**38**, pp. 75–96). Data on reliability and validity are included in the second report (**36**, pp. 21–45). It was called the "mosaic" test because it includes a wide variety of items.

the median interval. In the urban group the median interval was 24 months, in the rural 23 months. On the intelligence test, the children with long intersib intervals obtained significantly higher means, these differences persisting within each of the five occupational categories into which the sample was subdivided. With long intersib intervals, the scores approximated those of only children; with short intersib intervals, they approximated the scores obtained by three-child sibships.

Test scores were also available for 750 twins (**36**, Ch. 3; **90**). In this survey, family size averaged higher for twins than for the total sample, even when the twins were counted as a single birth. Analysis of paternal occupation showed a greater proportion of farmers in the twin sample than in the total sample. With regard to all other occupational classes, however, the twin sample was superior to the total sample. Thus a higher percentage of twins had fathers in the managerial and professional classes, and a lower percentage in the clerical and manual labor categories than was the case in the total sample. On the whole, then, the findings pertaining to occupation are similar to those of the Scottish survey, in which paternal occupation was somewhat superior for twins than for singletons.

The test scores of the twins were again inferior to those of the total sample. This difference was found at each age, and remained when family size and occupational level were controlled. In the case of families of four or more children, however, the twin inferiority was negligible and inconsistent, being most pronounced in the case of two- and three-sib families. Moreover, among the two-sib families, the twin scores were very similar to those of short-interval sibs, as defined in the earlier part of the study.

Other similar surveys have been conducted on a less extensive scale in several countries. In England, there have been a number of attempts to test complete age samples of children within restricted areas. Fraser Roberts and his co-workers (**77**, **79**) administered the Otis test to all children born during a four-year period and living within the city of Bath; a subsample was also individually tested with the Stanford-Binet. The Cattell Culture-Free Test was used by Cattell (**18**, **19**) with all 10-year-old children in Leicester, a typical industrial city, and Devonshire, an "unspoilt rural area." Sutherland and Thomson (**89**) applied a specially developed group intelligence test to all 11-year-old school children in the Isle of Wight. Burt (**14**) reports the results obtained with the elementary school population of one of the London boroughs in the course of his standardization of the Binet scale. The Otis test scores of 393 10-year-olds in Sheffield were analyzed by Bradford (**12**). All of these studies confirm the findings of the more extensive surveys regarding the decline in mean intelligence test score with increasing size of sibship. When correlations were computed between test score and family size, the coefficients were uniformly negative and clustered in the .20's and low .30's.

A similar inverse relation between intelligence and family size is reported by Papavassiliou (**72**), who gave a Greek adaptation of the Stanford-Binet to 349 children in Athens. Heinen (**37**) found a tendency for family size to be inversely related to school grades and other indices of academic ability in several large samples of German school children. In another survey of German school

children, Eydt (**34**) reported more retardation and behavior problems among children from large families.

Several American studies have yielded similar results. Lentz (**54**) obtained a correlation of −.304 between family size and intelligence test scores among 4,330 urban school children in five states. Within separate cities, however, the correlations varied from −.065 to −.411. This finding is not surprising in view of the fact that both the tests and the examiners varied from school to school. Chapman and Wiggins (**24**) reported a correlation of −.33 between National Intelligence Test scores and size of family among 650 school children. When the correlation was recomputed for children with native-born and with foreign-born parents, it fell to −.22 and −.24, respectively. The authors point out that most of the children came from completed families, since the youngest child in any sibship included in the study was 5 years old.

In a group of 1,140 8th grade school children in Oakland, California, Burks and Jones (**13**) found a correlation of −.22 between Terman Group Test and total number of siblings (including deceased siblings). This correlation rose to −.31 when corrected for restricted range in both IQ and sibship size. In order to include only completed families, the analysis was limited to sibships in which the youngest child was at least 9 years old. A correlation of −.31 between family size and Terman-McNemar intelligence test scores was obtained by Damrin (**26**) in a group of 156 high school girls.

Hirsch (**40**) found a consistent decrease in mean Otis IQ with increasing family size in a group of 214 Tennessee school children. The mean dropped from 115.8 for one-child families to 96.6 for eight-child families. Among still larger families, the mean varied irregularly, dropping to 82.0 for twelve-child families. In a study of more than 20,000 retarded children in the Massachusetts public schools, Dayton (**27**) noted a decrease in mean IQ from 73.2 for one-child families to 68.9 for families of ten or more children.

A more thorough analysis of the problem is provided in a study by Conrad and Jones (**25**), who tested representative samples of the rural population of three New England states. The children were examined with the Stanford-Binet, and the adults with the Army Alpha. Family size was investigated in relation to intelligence test score of mother and father, education of mother and father, a social status index of the family, and mean intelligence test score of the sibship. In uncompleted families, all of these measures correlated negatively with family size, although only the correlations with father's intelligence and with social status reached statistical significance. In the completed families, on the other hand, all correlations were positive, but none was statistically significant. Further analyses showed that the families of lower intelligence began to have children at an earlier age, although their total child-bearing period was no longer than in the more intelligent families. These conclusions, of course, must be restricted to the particular population under investigation.

Similar results were obtained, however, in a study of California families reported by Willoughby (**108**). On the basis of parental ages, the author concluded that most of the families were completed. Parental intelligence was determined from the combined score on six verbal and five nonverbal tests. Number of children correlated −.11 with the

test scores of 108 mothers and +.06 with the scores of 87 fathers, neither correlation being statistically significant.

Another study in which intelligence test scores of parents were employed is that of Willoughby and Coogan (**109**). Follow-up data were gathered on 373 persons who had graduated from high school 12 years earlier and who had taken an intelligence test while still in school. No relationship was found between the original test scores and age at marriage, age at birth of first child, or number of offspring at the time of the survey. It is also interesting to note that within the same group a low but significant negative correlation was found between intelligence test score and the number of siblings of the subjects themselves. Moreover, the subsequent education and the occupational level of the subjects revealed an inverse relation to number of offspring. Although this study is inconclusive because of the small and selected sample and because of the incompleteness of the families, it nevertheless represents a promising approach to the problem.[4] Its chief merits lie in the fact that intelligence of parents was investigated directly and in the fact that such intelligence was measured prior to educational and vocational differentiation.

An investigation currently in progress provides further information on the relationship between intelligence of parents and number of offspring.[5] The original data for this study were gathered by E. Lowell Kelly as part of a broader project involving the testing, interviewing, and subsequent 20-year follow-up of 300 engaged couples living in New England (**46**). The data pertaining to fertility, based on 216 couples, are being processed by Charles F. Westoff and Elliot G. Mishler of the Office of Population Research, Princeton University. At the time of the initial testing, the mean age of the men was 26.6 years and that of the women 24.6, most of the subjects falling within the age limits of 21 and 30. Since data on number of offspring were obtained 20 years later, it can be assumed that most of the families were completed. Educationally, the group represented a superior sample, more than half of the men and approximately a third of the women being college graduates at the time of initial testing. The correlation between initial Otis Intelligence Test scores and subsequent number of live births was .19 for the 216 men and .17 for the 216 women, both being significant at approximately the .01 level. These correlations thus suggest a slight tendency for the brighter parents to have *larger* families. It is noteworthy that none of the studies in which parental intelligence was correlated with number of offspring has so far yielded the negative correlation customarily found when child intelligence is correlated with size of sibship.

The relationship between fertility and intelligence has also been investigated within certain special groups. Moshinsky (**68**) analyzed the test scores of 10,159 English children enrolled in different types of schools. In the relatively heterogeneous elementary school group, a correlation of −.23 was found between intelligence and family size. Within more advanced, fee-paying schools, the correlations were negligible. Both Moshinsky and Blackburn (**8**) cite the latter finding in support of cer-

[4] Cf. critique of this study by Fraser Roberts (**78**), but note that one of the criticisms is based upon a confusion between number of siblings and number of offspring.

[5] Personal communication from Dr. Charles F. Westoff, November 4, 1955. Cf. also (**67**).

tain hypotheses regarding the relationship between intelligence test performance and family size. As Burt (**15**) points out, however, such a result is to be expected because of the greater intellectual homogeneity of the children in the more advanced schools.

Hutton (**42**) reports an insignificant correlation of −.05 between IQ and number of siblings in the case of boys attending a British secondary grammar school. The group is described as falling within the upper half of the general population in intelligence. A questionnaire survey of alumni of the same school indicated that the highly selected scholarship holders as a group tended to have a smaller number of offspring than the general population of Great Britain. But among the more recent graduates the reverse was true, the scholars showing a higher replacement rate than the generality.

A number of studies on college students, in both Great Britain and the United States, have yielded very low and usually insignificant correlations between intelligence test scores and number of siblings (**39, 64, 65, 85, 102, 103**). This lack of correlation is understandable in the light of the highly selected nature of such samples. Not only are college students selected intellectually; they are also selected with respect to the constellation of psychological and socioeconomic factors which determine who goes to college. Some of these factors may themselves be directly related to family size.

In their study of California public school children with Stanford-Binet IQ's of 140 or higher, Terman and his co-workers found a correlation of −.271 between IQ and size of sibship within 91 completed families (**91**). Through the subsequent follow-ups of the entire gifted group, additional data have been gathered on the completed family size of nearly all of the original cases. These data, however, have not yet been analyzed. The forthcoming fifth volume of *Genetic Studies of Genius* will contain extensive data on the relation between intelligence and number of offspring of the gifted subjects themselves.[6]

In connection with a slum clearance survey, Dawson (**28**) administered the Stanford-Binet to 1,239 Glasgow children, aged 3 to 14. Most of the children were tested just before their families moved to a new housing development. The correlation between these IQ's and the number of living sibs was −.12; within a subgroup of 140 completed families, the correlation was −.10. The corresponding correlations with total number of births (including stillbirths) were −.20 and −.30, respectively. O'Hanlon (**71**) reports the results of retesting 293 of these children eight years later. This group yielded a correlation of −.137 between IQ and number of living sibs, the correlation rising to −.291 in a subgroup of 28 completed families. The corresponding correlations with total number of births were −.207 and −.413. The results of both Dawson and O'Hanlon illustrate the fact that differential fertility is partly offset by a differential death rate during childhood.

In an early study, Pearson and Moul (**73**) found no significant correlation between size of sibship and teachers' ratings of intelligence among 1,202 Jewish school children in London. The children were largely of foreign-born parentage and came from relatively homogeneous and low socioeconomic levels. Fertility was uniformly high in the group as

[6] Personal communication from Dr. Lewis M. Terman, April 28, 1955.

a whole. Sutherland (**87**) reports correlations of −.129 and −.126 between IQ and family size in two samples of children of British coal miners totaling 3,096 cases. These correlations are consistent with the findings of other investigators insofar as the relationship is somewhat lower in the poorest socioeconomic levels than in the intermediate levels.

In a study of 354 children in a Jewish orphanage in New York City, Locke and Goldstein (**56**) found a correlation of −.13 between Stanford-Binet IQ and size of sibship. When both mother's age and order of birth were partialled out, this correlation rose to −.24. It is rather difficult, however, to interpret this partial correlation. For comparability with correlations obtained in other studies, moreover, the original correlation of −.13 would seem preferable.

Sutherland (**88**) analyzed the IQ's of two samples of fatherless children. The first consisted of 123 12- to 14-year-old Glasgow school children whose fathers had died when the children were under one year of age. The second sample comprised 724 Yorkshire school children aged 11 to 13 who were also fatherless but not necessarily from infancy. Each sample was compared with a control group of children both of whose parents were living. The mean IQ of the fatherless children was lower than that of control children with the same number of siblings. The correlations between IQ and family size were sigficantly negative, but lower among fatherless children than among the controls. The respective correlations were −.188 and −.26 in the Edinburgh samples, and −.19 and −.23 in the Yorkshire samples.

In discussing these results, Sutherland writes, "The tendency for the brighter children to come from the smaller families has been reduced by the presence of small families whose natural increase has been prevented by death and whose children would normally have had their place among the larger families" (**88**, p. 168). A similar explanation may account for the lower IQ's of the fatherless children within each sibship size. The test performance of such children resembled that of children in larger families, to which they would have belonged if family growth had not been cut off by the father's death. This type of study represents one step toward the analysis of the complex factors determining the correlation between intelligence and family size. Its findings suggest that it is not size of sibship per se but other factors associated with family size within a given culture which produce the obtained differentials in intellectual level.

Follow-up Studies with Intelligence Tests

On the basis of the negative correlation between intelligence test scores and family size, several writers predicted a gradual decline in the intellectual level of the population (**14**, **18**, **19**, **20**, **54**, **77**, **79**, **92**, **100**). The estimated drop varied from about 2 to 4 IQ points per generation. A direct test of such a predicted decline is provided by follow-up surveys in which comparable samples have been tested under similar conditions after a lapse of several years. The most extensive and best controlled data on this question are to be found in the previously described Scottish surveys (**81**, **82**, **83**, **93**, **94**). Rather than showing the predicted decline, however, the scores revealed a small but significant improvement from the 1932 to the 1947 surveys.

Other similar follow-ups have been reported by Burt (**14**), Cattell (**23**),

and Emmett (**33**). In comparison to the Scottish surveys, these studies covered somewhat shorter intervals and used smaller samples of questionable comparability. When testing is conducted within a limited area, such as a single county, city, or borough, successive samples may differ owing to selective migration, deterioration of neighborhoods, slum clearance, changes in proportion of children institutionalized or attending private schools, and other similar factors. Moreover, differences in thoroughness of sampling procedures in successive surveys may introduce differential selection. For example, subjects in lower socioeconomic levels and those who are physically less fit are more likely to be excluded when sampling coverage is less adequate. Despite certain local variations in results, however, these studies also failed to substantiate the expected decline in test score.

Reference may likewise be made to certain relevant data gathered incidentally, as by-products from other types of testing programs. Thus an analysis of the intelligence-test performance of American soldiers in World Wars I and II indicated that the level of performance had improved to such an extent that the median score of the later sample corresponded to the 83rd percentile of the earlier sample (**95**). This rise in score paralleled an increase in amount of education, the mean being at the 8th grade for the first sample and at the 10th grade for the second. It is also noteworthy that a survey of the intelligence test performance of American high school students over a 20-year period suggested that this, too, had improved, despite the marked increase in the proportion of students enrolled in high school (**35**). Since a larger proportion of the total population was attending high school at the end than at the beginning of this 20-year period, a decrease in mean score would be expected, unless the total population had improved sufficiently to counteract such a drop.

When educational and other environmental conditions within a given community improve conspicuously, larger rises in test scores may be observed over even shorter intervals. Thus a 10-year follow-up conducted in a relatively isolated mountainous region of eastern Tennessee showed a 10-point rise in median IQ (**105**). During this interval, the socioeconomic and educational conditions of this region had greatly improved. The second sample tested in this survey was closely comparable to the first, the children coming largely from the same families in the two cases.

Similar results are to be found in the previously cited slum survey conducted by Dawson (**29**). A subgroup of 289 4- to 9-year-old children who had been tested with the Stanford-Binet at the time of their transfer from the slum area was retested after 12 to 18 months of residence in improved housing. A control group of 56 cases who had not been transferred from the slum area was also retested over the same interval. The slum clearance group showed a significant mean rise of 1.5 IQ points, while the control group showed no significant change. To be sure, the mean IQ rise in the clearance group, though significant, was small. But it should be borne in mind that the retest was conducted over a relatively short period of time. Moreover, it is noteworthy that the improvement of the slum clearance group was also manifested on tests of arithmetic and reading, and persisted when comparisons were made separately for boys and girls in all three variables. In the control group, on the other hand, about half of these comparisons yielded positive changes and half negative changes.

That intelligence test performance is susceptible to improvement as a result of education has been repeatedly demonstrated by a variety of methods (**4**, Ch. 8). Of special interest are two investigations in which persons who had been tested following a uniform period of universal education were retested after 10 years in one case and after 20 years in the other (**41**, **58**). In the intervening years, members of these groups had received varying amounts of further education. Both studies report significant relationships between amount of subsequent schooling and retest performance, when initial scores are held constant.

One of the hypotheses proposed to account for the rise in mean score found in the Scottish surveys and similar follow-ups is that of test sophistication. Thus it is argued that the subjects tested in the later samples had the advantage of greater familiarity with psychological tests. With regard to the Scottish surveys themselves, a special analysis of the scores obtained by children who had and those who had not been previously examined with a similar test showed little effect of such test experience (**83**, pp. 121–124). In a later and more extensive study of the same question, two subgroups within the 1947 Scottish sample were compared. One was taken from Education Authorities in which only 2.5% of the children were reported to have had previous test experience within the year of the survey; the other included Educational Authorities in which 67% of the children had had such experience. The mean rise in score from 1932 to 1947 in the two subgroups was 1.5 and 3.4 points, respectively, both being significant at the .01 level.[7]

[7] Personal communication from Mr. James Maxwell, University of Edinburgh, May 17, 1955.

Fundamentally, the question of test sophistication and the related question of coaching must be answered in terms of breadth of improvement (cf. **1**; **2**, pp. 52–56). Any influence which is restricted to the test performance itself and does not correspondingly affect the criterion behavior which the test is designed to predict would of course reduce the validity of the test. But the broad social and educational changes which have occurred in the time intervals under consideration can be expected to affect the individual's over-all intellectual development, rather than being limited to the particular behavior samples covered by specific tests.

Carrying the argument a bit further, some writers have argued that "true, innate" intelligence may still have declined, despite the rise in intelligence test performance brought about by improved environmental conditions. Some have gone so far as to suggest that if test scores show improvement under such conditions, then the tests are at fault. This is analogous to arguing that the "true, innate" height of the population has declined as a result of the negative correlation between height and family size, and that the observed rise in mean height is illusory. Moreover, it could then be argued that we should devise a "culture-free" meterstick to measure innate height freed from environmental influences. Such reasoning does not make scientific sense. When the psychologist speaks of intelligence, he refers to certain properties of observable behavior. Such behavior is by its very nature susceptible to environmental influences. The same influences which bring about a rise in test score also cause an improvement in the quality of the individual's general intellectual functioning (cf. **31**, **75**, **100**).

Penrose (**74**, **75**, **76**) has argued

that, even from a genetic point of view, the negative correlation between intelligence test scores and size of sibship need not imply a decline in the intellectual level of the population. Taking into account assortative mating and the infertility of low-grade mental defectives, he presents a simplified genetic model by which intellectual level remains constant despite the existing fertility differentials. Essentially, such a situation results from the fact that some parents of average or borderline intelligence do have above-average children. In a much earlier article by Willoughby and Goodrie (**110**), a similar argument is presented. Beginning with a correlation of −.17 between intelligence and family size, the authors set up a genetic model based upon the combined contribution of five gene pairs. They then trace the effects mathematically through five generations, using two marital coefficients, .37 and .58. With the former coefficient, a decline in mean intelligence results. With the latter, which is much closer to empirically obtained husband-wife correlations in intelligence test scores, there is actually a slight rise in mean intelligence.

Methodological Problems

Investigations on the relationship between family size and intelligence present a number of special methodological problems with regard to sampling and statistical analysis. A source of difficulty in several studies is the inclusion of *incomplete families* in the sample. This factor may operate in a number of different and not entirely predictable ways. All children from such incomplete families are, of course, classified as coming from somewhat smaller families than will eventually be the case. This may have the effect of classifying lower test-scorers (from the ultimately larger families) into a category of higher test-scorers (from smaller families). Such a procedure would thus reduce the obtained differences in test scores between the various family-size categories. On the other hand, when a negative correlation is found between intelligence and family size in a sample of incomplete families, it may simply reflect the tendency for persons of lower intelligence to begin having children earlier. Should such persons also stop having children earlier than the more intelligent members of the community, the correlation would disappear when completed families are studied. In the previously reported investigation by Conrad and Jones (**25**) in rural New England communities, such was actually the case.

It should be noted, of course, that if the less intelligent parents begin having children earlier than the more intelligent parents, they will contribute more individuals to the population in the long run, through more frequent generations. For example, if each successive generation begins to bear children at age 20, there will be five generations per century; while if child-bearing begins at age 25, there will be only four generations. Nevertheless, the effect upon the intellectual level of the population would be less if the duller parents merely began to have children earlier than it would be if the duller also had larger completed families.

Age of parents is a factor which needs to be considered in its own right. The smaller families in a sample may include a relatively large proportion of incomplete families of younger parents. Such parents are likely to be better educated than the older parents of the completed families, owing to the progressive rise in educational level of the population. The better educated parents would in turn provide a more favorable en-

vironment for their offspring. Thus when incomplete families are included in the sample, the differences in parental age, coupled with the rising level of general education, may account at least in part for the negative correlation found between family size and intelligence.

A second major problem pertains to *selective factors*. Since the correlations in question are generally quite low, the operation of selective factors, however slight, may produce a completely spurious result. An example of such a subtle selective factor is given in Shuey's study of students in an American women's college (**85**). Within the entire group of 2,261 entering students for whom test scores were available, a significant inverse relation was found between test scores and number of siblings. This relationship disappeared, however, when the records of those whose older sisters had attended the same college were excluded. Presumably somewhat lower entrance requirements had been applied to those whose siblings had previously matriculated. Such students would also be likely to come from larger families.

Nor are more general samples free from selective factors. Thus in a sample of children between, let us say, the ages of 8 and 14, individual large families will appear repeatedly within the sample, the number of reappearances being directly proportional to the size of the family. A family containing five school-age children, for example, would appear five times in the sample. Any chance conditions characterizing individual large families would thus be spuriously magnified in the sample. Nor would ordinary sampling statistics be strictly applicable to such samples. It is partly to avoid such difficulties that single-age samples, such as 11-year-olds, have been utilized.

Even carefully chosen single-age samples, however, are not entirely free from selective bias. In such a sample, for example, large families will be overrepresented, since by chance a child of any specified age is more likely to be found in a large family than in a small one. Hence the sample is not representative of existing families, nor of the parental population. With regard to the child population, it can be argued that large families should be overrepresented, since they contribute more members to the population. Nevertheless, certain types of comparisons may be distorted through the use of a single-age sample. The reports of the Scottish survey contain a discussion of the possible effects of such selective biases upon the analysis of birth order (**82**) and maternal age (**83**) in relation to child's intelligence.

Parenthetically, it may be added that studies on the relation of birth order to intellectual and other psychological characteristics have frequently yielded ambiguous and inconsistent results because of the failure to take family size into account. Let us suppose that one group consists entirely of two-child families, while a second group contains only four-child families. Obviously, in the first group 50 per cent of the children will be first-born, while in the second there will be only 25 per cent. In the Scottish survey (**82**), the child's position in his sibship was recorded as ¼, ⅖, etc., signifying the first-born in a sibship of four, the second-born in a sibship of five, and so forth. Comparisons were then made within each family size. Under these conditions, the first-born and last-born scored higher than the intermediate sibs. But even these results are inconclusive because of the previously mentioned sampling biases.

The computation of *predicted intellectual decline* involves questionable assumptions. The usual pro-

cedure is based essentially upon the comparison of two means obtained from the same set of scores (cf., e.g., **14**, **19**). The mean of the subjects' scores can of course be computed directly. The mean of the next generation is estimated by multiplying the frequency in each class-interval of the IQ distribution by mean family size in that interval and using these new frequencies to find the mean. Similarly, the mean score of the parents of the present subjects is estimated by dividing the original frequencies by family size and multiplying by two.

Such a procedure is based upon the assumption of a perfect parent-child correlation in both intelligence test score and family size. In neither case is the assumption consistent with available information. Parent-child correlations in intelligence test scores are generally in the .50's (cf. **4**, pp. 318–320). With regard to fertility, parent-child discrepancies may arise from various sources, including the decrease in fertility differentials among socioeconomic classes and the social mobility of individuals. Available data indicate that when persons move from one culture or subculture to another, their fertility tends to be intermediate between that of the group from which they have come and that of the group which they join.

Another noteworthy methodological point pertains to the definition of *mean family size*. When size of sibship is recorded for each child and these sizes are averaged, the mean family size *per child* is obtained. On the other hand, when number of children in each family is recorded, as in a census survey, the average of these numbers gives mean family size *per family*. Thus if we have a sample of 5 families consisting of 1, 2, 3, 4, and 5 children, respectively, the median family has 3 children. But the median child, out of the total 15 in the sample, belongs to a family of 4. The corresponding means are 3 for families and 3.67 for children. The latter is the contraharmonic mean discussed by Jaspen (**44**, **45**) and bears a direct arithmetic relation to the former. If X represents number of children in each family and N represents number of families, the mean per family is $\Sigma X/N$ and the mean per child is $\Sigma X^2/\Sigma X$. In making comparisons between mean family sizes of populations or segments of a population, it is obviously essential to insure that the same type of mean is employed. This point is especially likely to arise when results obtained with a child sample are compared to national census figures, since the former are usually expressed as mean per child, the latter as mean per family.

Attention should likewise be given to certain problems dealing with the choice and application of psychological tests in studies on fertility and intelligence. An intrinsic feature of the design of the Scottish surveys involved the *transmutation of scores* from a group test, administered to the entire sample, to Stanford-Binet IQ's obtained from a cross section of the larger sample. It was apparently felt that the scores of the entire sample would gain in meaningfulness if put in terms of such IQ's. Apart from the wider familiarity of the IQ, however, it is difficult to see what such a conversion accomplishes. Moreover, a number of specific objections can be raised against this procedure. Since different intelligence tests measure a somewhat different combination of functions, individuals cannot be expected to retain the same relative standing on such tests. Intelligence tests vary not only in score scale and in normative population,

but also in content. The reliability and validity of the converted scores depend upon the properties of the test which was actually administered, rather than upon the properties of the test into which the scores were converted. For example, if an individual test is more precise because of its better control of rapport and motivation, longer testing time, or any other favorable conditions, these characteristics are not transferred to the group test by transmutation of scores. Nor is the use of the traditional ratio IQ advisable for such purposes. Standard scores, expressed as deviation IQ's or in some other convenient form, have far more to recommend them for precise measurement.

Finally, the exclusive use of *global measures of intelligence* in such studies may be questioned. In the light of the findings of factor analysis, it would appear desirable to investigate correlations with family size, as well as temporal changes in population means, in terms of more nearly homogeneous intellectual functions. It would be of considerable interest to discover, for example, whether verbal aptitudes are more highly correlated with family size than spatial or numerical aptitudes, and whether the rise in mean score found in follow-up surveys is uniformly high in all aspects of intelligence.

Analysis of Causal Relations

Even with similar results, different investigators have offered varied interpretations of the correlation between intelligence and family size. It is especially important in this connection to differentiate between plausible hypotheses and empirically established facts, and to ferret out tacit assumptions which may underlie certain interpretations. For example, if intelligence is defined in terms of hereditary predispositions, it becomes logically impossible to investigate whether or not certain individual differences in intelligence may be related to gene constitution. It is more consistent with scientific method to employ definitions which do not prejudge the data, but which permit the maximum freedom of interpretation.

We may, of course, define the word "intelligence" to mean anything we choose. All studies on the relation of intelligence to family size, however, have employed psychological test scores as their index of intelligence. It is therefore evident that our interpretations must take such test scores as their point of departure. When intelligence is defined in such terms, we may recognize three distinct etiological mechanisms whereby the obtained correlations between intelligence and family size might result (cf. **3**, **59**). It is understood, of course, that the actual causal relations may involve any combination of two or all three of these mechanisms.

First, there may be *inherited structural factors* (neural, glandular, etc.) which serve as constraints, reducing adaptability of behavior and limiting the sort of intellectual development measured by current intelligence tests. The less able parents would thus transmit their hereditary limitations to their offspring. The obtained correlations would then result from the fact that, within a given culture or subculture, persons with inferior heredity tended to have more offspring.

A second explanation attributes individual differences in children's abilities to *psychological differences in the environments* provided by parents of varying intellectual levels. In this case, the correlation between family size and intelligence of off-

spring would again result from a tendency for the less intelligent parents to have more children, but heredity would not be involved. Differences in intellectual level among the parents, as well as among the offspring, could thus have resulted from environmental factors.

In this connection, mention should also be made of the point of view represented by Vernon (**101**). In a discussion of the use of intelligence test scores in population studies, he urges the need for caution in interpreting such scores because of the effect of differences in cultural patterns upon test performance. Such cultural differences, however, may affect not only test scores but also the individual's over-all intellectual development. The real question pertains not to inadequacies of available measuring instruments but rather to the interpretation of the existing behavioral differences.

It may be added that the individual's intellectual development may be influenced, not only by the nature and extent of direct intellectual stimulation offered by his home, but also by emotional and motivational factors deriving from the "social climate" of the home. There is a growing body of data pertaining to differences in child-rearing practices in various socioeconomic levels and other subcultures (**4**, Ch. 23 and p. 734). Some of these differences, such as the degree to which verbalization and exploratory behavior are encouraged or discouraged, are likely to influence the child's intellectual development.

A third possible interpretation of the obtained correlations between intelligence and fertility is based upon *size of family* itself as a causal factor. For example, a larger family—at least in certain socioeconomic levels—would reduce the per capita funds available for education, recreation, suitable housing, proper food, medical attention, and other environmental requisites. From a psychological viewpoint, another important factor is the degree of adult contact provided in families of different sizes. Available evidence suggests, for example, that such contact may be the most important single factor in linguistic development (**4**, pp. 335–339; **62**). And it is well known that verbal ability plays a major role both in educational progress and in intelligence test performance. Whatever the specific factors involved, the causal mechanism under consideration is independent of the intellectual level of parents. A crucial test of this hypothesis would thus be found in the correlation between size of sibship and intelligence of offspring, when intellectual level of parents is held constant.

One additional point may be considered. Insofar as family size itself may be a factor in intellectual development, the general decrease in family size in the period covered by most follow-up surveys is noteworthy. Quite apart from fertility differentials, such a decrease may have operated as one factor in the observed rise in intellectual level.

It is evident that the three hypotheses differ significantly in both their theoretical and practical implications. On the basis of existing data, it is impossible to choose among them or to determine the relative contributions of each of the three types of influences. The complex interaction of many variables makes analysis of causal relations difficult in this area. A few of the previously cited results, however, have a bearing upon causal interpretations. Thus Sutherland's study of fatherless children suggested that family size per se is insufficient to account for the obtained

correlations between family size and intelligence (**88**). On the other hand, the findings on twins in both the Scottish (**83**) and French (**36, 90**) surveys are consistent with interpretations in terms of adult contact. The results of Tabah and Sutter (**36, 90**) on intersib interval likewise seem to support such an explanation. The fact that studies of parents have so far failed to demonstrate a negative correlation between parental intelligence and number of offspring is also in line with the third hypothesis outlined above.

Vernon (**101**) recommends the study of orphans and foster children whose true parents' families differ in size. Such an approach would undoubtedly provide crucial data on the first hypothesis, which concerns the contribution of hereditary factors to the obtained correlation between intelligence and family size. This type of investigation, however, would be handicapped by a number of practical obstacles, such as the limitation in number and range of available cases. Moreover, family size of orphans and foster children would be artificially restricted by such conditions as death of one or both parents and illegitimacy.

A study specifically designed to test the role of parental contact in the association of intelligence and family size is reported by Nisbet (**69, 70**). Three types of data were obtained on several thousand Aberdeen school children. First, in the effort to gauge the contribution of language development to the correlation between family size and performance on the ordinary verbal type of intelligence test, the latter correlation was found with language test scores partialled out. This partial correlation was only −.04. When family size and language test score were correlated, with intelligence test score partialled out, a higher correlation (−.11) was obtained. Secondly, verbal intelligence tests yielded significantly higher correlations with family size than did such nonverbal tests as Raven's Progressive Matrices. Thirdly, the correlation between intelligence test scores and family size rose with age. The latter finding was obtained in both a cross-sectional study of the large sample and a longitudinal study of a small sample. From the results of these three approaches, Nisbet concluded that part of the negative correlation between family size and intelligence may be attributed to the effect of sibship size upon verbal development, which in turn influences intellectual development.

In a later study on 288 adult women, Scott and Nisbet (**80**) again found that the negative correlation between size of sibship and intelligence test score is lower for nonverbal than for verbal tests. Such results corroborate Nisbet's findings on the child sample described above, and suggest that the environmental factors associated with family size may affect test performance well into adulthood.

It should be noted that the psychological effect of family size upon intellectual development may be curvilinear rather than rectilinear. Moreover, the direction in which increasing family size influences intellectual development may itself differ with concomitant psychological and social circumstances. Attention should likewise be given to the possible effects of family size upon nonintellectual characteristics, such as socialization, cooperation, emotional security, leadership, and other aspects of interpersonal relations. Available research on this question has yielded conflicting and ambiguous results (**9, 10, 11, 26, 27, 32, 55, 57, 86**). Some

studies provide evidence that certain desirable emotional characteristics, as well as acceptance by associates, may be positively related to membership in relatively large families (**9**, **11**, **27**). Most of the results, however, are complicated by the correlation between indices of social adjustment, on the one hand, and both intellectual and socioeconomic level, on the other. Differences in family size among ethnic subgroups within a sample are another source of confusion in such data.

In recognition of the methodological and interpretative complexities of the problem, there is a growing interest in the design of more nearly definitive investigations on the relationship between intellectual level and fertility.[8] Ideally, such investigations should begin with the testing of young people prior to their educational and vocational differentiation, i.e., after all have completed a uniform period of required schooling. Preferably the test should consist of a differential aptitude battery yielding a profile of scores rather than a single global measure. The subjects should be followed up until all or nearly all of their families are completed. Age of both parents at the birth of their first and last child should be recorded. Data should also be kept on deaths, unmarried persons, and childless marriages. Information should likewise be gathered regarding occupation, income level, and amount of subsequent education for each member of the group. It would also be of interest to obtain indices of social mobility, such as changes in occupational, educational, or income level within the subject's own life, as well as differences between his status and that of his parents.

Additional questions can be answered if test scores of the children of these persons are studied; but many problems can be solved even prior to this step. From a practical viewpoint, such a program is not unrealistic, especially in nations where uniform psychological tests are administered in the school system and where detailed census data are regularly collected. From a theoretical standpoint, this approach would help to separate the many interrelated variables which are now intricately intertwined, and should thus bring us closer to a causal interpretation of the empirically observed relationship between intelligence and family size.

[8] Cf., e.g., discussions by Anastasi (**3**), Cattell (**21**), and Vernon (**101**), as well as the Summary Report of Meeting 25, World Population Conference (**97**).

REFERENCES

1. Anastasi, Anne. The concept of validity in the interpretation of test scores. *Educ. psychol. Measmt*, 1950, **10**, 67–78.
2. Anastasi, Anne. *Psychological testing*. New York: Macmillan, 1954.
3. Anastasi, Anne. Tested intelligence and family size: methodological and interpretive problems. *Eugen. Quart.*, 1954, **1**, 155–160.
4. Anastasi, Anne, & Foley, J. P., Jr. *Differential psychology*. (Rev. Ed.) New York: Macmillan, 1949.
5. Anonymous. Fertility and educational level of parents. *Stat. Bull. Metropl. Life Insur.*, 1945, **26**, No. 11, 6–7.
6. Baltzell, E. D. Social mobility and fertility within an elite group. *Milbank Mem. Fund Quart.*, 1953, **31**, 411–420.
7. Berent, J. Fertility and social mobility. *Popul. Stud.*, 1952, **5**, 244–260.
8. Blackburn, J. Family size, intelligence score, and social class. *Popul. Stud.*, 1947, **1**, 165–176.
9. Bonney, M. E. A study of the relation of intelligence, family size, and sex differences with mutual friendships in the primary grades. *Child Developm.*, 1942, **13**, 79–100.
10. Bonney, M. E. Relationships between

social success, family size, socio-economic home background, and intelligence among school children in grades III to V. *Sociometry*, 1944, **7**, 26–39.

11. BOSSARD, J. H. S., & BOLL, ELEANOR S. Security in the large family. *Ment. Hyg., N. Y.*, 1954, **38**, 529–544.
12. BRADFORD, E. J. G. Can present scholastic standards be maintained. *Forum Educ.*, 1925, **3**, 186–198.
13. BURKS, BARBARA S., & JONES, H. E. A study of differential fertility in two California cities. *Hum. Biol.*, 1935, **7**, 539–554.
14. BURT, C. *Intelligence and fertility.* London: Hamilton, 1946.
15. BURT, C. Family size, intelligence, and social class. *Popul. Stud.*, 1947, **1**, 177–186.
16. CARTER, C. O. Differential fertility in 1951. *Eugen. Rev.*, 1953, **45**, 101–103.
17. CARTER, C. O. Differential fertility and intelligence. *Bull. World Fed. ment. Hlth*, 1954, **6** (2), 101–103.
18. CATTELL, R. B. Is national intelligence declining? *Eugen. Rev.*, 1936, **28**, 181–203.
19. CATTELL, R. B. *The fight for our national intelligence.* London: King, 1937.
20. CATTELL, R. B. A study of the national reserves of intelligence. *Hum. Fact.*, 1937, **11**, 127–137.
21. CATTELL, R. B. Some further relations between intelligence, fertility, and socio-economic factors. *Eugen. Rev.*, 1937, **29**, 171–179.
22. CATTELL, R. B. Intelligence and fertility. *Eugen. Rev.*, 1945, **36**, 126–127.
23. CATTELL, R. B. The fate of national intelligence: test of a thirteen-year prediction. *Eugen. Rev.*, 1951, **42**, 136–148.
24. CHAPMAN, J. C., & WIGGINS, D. M. Relation of family size to intelligence of offspring and socio-economic status of family. *J. genet. Psychol.*, 1925, **32**, 414–421.
25. CONRAD, H. S., & JONES, H. E. A field study of the differential birth rate. *J. Amer. Statist. Ass.*, 1932, **27**, 153–159.
26. DAMRIN, DORA E. Family size and sibling age, sex, and position as related to certain aspects of adjustment. *J. soc. Psychol.*, 1949, **29**, 93–102.
27. DAYTON, N. A. Influence of size of family upon the characteristics of the mentally deficient. Survey of 20,473 retarded children in the public schools of Massachusetts. *Amer. J. Psychiat.*, 1935, **91**, 799–832.
28. DAWSON, S. Intelligence and fertility. *Brit. J. Psychol.*, 1932, **23**, 42–51.
29. DAWSON, S. Environmental influence on mentality. *Brit. J. Psychol.*, 1936, **27**, 129–134.
30. DINKEL, R. M. Occupation and fertility in the United States. *Amer. sociol. Rev.*, 1952, **17**, 178–183.
31. DUNCAN, O. D. Is the intelligence of the general population declining? *Amer. sociol. Rev.*, 1952, **17**, 401–407.
32. EKSTRAND, RUTH. The variation of behavior problems with size of family. *Smith. Coll. Stud. Work*, 1931, **1**, 290–300.
33. EMMETT, W. G. The trend of intelligence in certain districts of England. *Popul. Stud.*, 1950, **3**, 324–337.
34. EYDT, A. Auslese und Ausmerze in der Volksschule. *Volk u. Rasse*, 1938, **13**, 344–351.
35. FINCH, F. H. Enrollment increases and changes in the mental level. *Appl. Psychol. Monogr.*, 1946, No. 10. Pp. 75.
36. GILLE, R., HENRY, L., TABAH, L., SUTTER, J., BERGUES, H., GIRARD, A., & BASTIDE, H. Le niveau intellectuel des enfants d'âge scolaire: la détermination des aptitudes; l'influence des facteurs constitutionnels, familiaux, et sociaux. *Inst. nat. d'études démographiques: Travaux et documents*, Cahier **23**, 1954. Pp. 294.
37. HEINEN, A. Schulleistungen, väterlicher Beruf, und Kinderzahl. *Z. mensch. Vererb.- u. Konst.-Lehre*, 1938, **21**, 599–608.
38. HEUYER, G., PIÉRON, H., PIÉRON, MME. H., & SAUVY, A. Le niveau intellectuel des enfants d'âge scolaire: une enquête nationale dans l'enseignement primaire. *Inst. nat. d'études démographiques: Travaux et documents*, Cahier **13**, 1950. Pp. 283.
39. HIMMELWEIT, HILDE T. Intelligence and size of family; their relationship in an adult group of high educational standard. *Eugen. Rev.*, 1948, **40**, 77–88.
40. HIRSCH, NATHANIEL, D. M. An experimental study upon three hundred school children over a six year period. *Genet. Psychol. Monogr.*, 1930, **7**, 493–549.
41. HUSÉN, T. The influence of schooling upon IQ. *Theoria*, 1951, **17**, 61–88.
42. HUTTON, K. Intelligence quotients and differential fertility. *Eugen Rev.*, 1953, **44**, 205–215.
43. JAFFE, A. J. Urbanization and fertility. *Amer. J. Sociol.*, 1942, **48**, 48–60.
44. JASPEN, N. A class of means. Paper

read at East. Psychol. Ass., New York, April, 1954.

45. JASPEN, N. Means and their standard errors. Paper read at Amer. Psychol. Ass., New York, September, 1954.
46. KELLY, E. L. Consistency of the adult personality over twenty years. *Amer. Psychologist*, 1955, **10**, 659–681.
47. KISER, C. V. *Group differences in urban fertility*. Baltimore: Williams & Wilkins, 1942.
48. KISER, C. V. Exploration of possibilities for new studies of factors affecting size of family. *Milbank Mem. Fund Quart.*, 1953, **31**, 456–479.
49. KISER, C. V. The Indianapolis fertility study—an example of planned observational research. *Publ. Opin. Quart.*, 1953–54, **17**, 496–510.
50. KISER, C. V., & WHELPTON, P. K. Social and psychological factors affecting fertility: II. Variations in the size of completed families of 6551 native-white couples in Indianapolis. *Milbank Mem. Fund Quart.*, 1944, **22**, 72–105.
51. KISER, C. V., & WHELPTON, P. K. Resumé of the Indianapolis study of social and psychological factors affecting fertility. *Popul. Stud.*, 1953, **7**, 95–110.
52. KUCZYNSKI, R. R. British demographers' opinions on fertility, 1660–1760. *Ann. Eugen.*, 1935, **6**, 139–171.
53. LEE, E. S., & LEE, ANNE S. The differential fertility of the American Negro. *Amer. sociol. Rev.*, 1952, **17**, 437–447.
54. LENTZ, T. F., JR. Relation of IQ to size of family. *J. educ. Psychol.*, 1927, **18**, 486–496.
55. LEVY, J. A quantitative study of behavior problems in relation to family constellations. *Amer. J. Psychiat.*, 1931, **10**, 637–654.
56. LOCKE, N. M., & GOLDSTEIN, H. The relation of birth order, age of mother, and size of family to intelligence. *J. Psychol.*, 1937, **3**, 89–96.
57. LOOMIS, C. P., BAKER, W. B., & PROCTOR, C. The size of the family as related to social success of children. *Sociometry*, 1949, **12**, 313–320.
58. LORGE, I. Schooling makes a difference. *Teachers Coll. Rec.*, 1945, **46**, 483–492.
59. LORIMER, F. Trends in capacity for intelligence. *Eugen. News*, 1952, **37**, 17–24.
60. LORIMER, F., *et al.*, *Culture and human fertility*. Paris: Unesco, 1954.
61. LORIMER, F., & OSBORN, F. *Dynamics of population*. New York: Macmillan, 1934.
62. MCCARTHY, DOROTHEA. Language development in children. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1954.
63. MACMEEKEN, AGNES M. *The intelligence of a representative group of Scottish children*. London: Univer. of London Press, 1939.
64. MAXWELL, ALINE, & HUESTIS, R. R. Student test-score rank and family size. *J. Hered.*, 1930, **21**, 211–215.
65. MAXWELL, J. Intelligence and family size of college students. *Eugen. Rev.*, 1951, **42**, 209–210.
66. MAXWELL, J. Intelligence, fertility, and the future: A report on the 1947 Scottish Mental Survey. *Eugen. Quart.*, 1954, **1**, 244–274.
67. MISHLER, E. G., WESTOFF, C. F., & KELLY, E. L. Some psychological correlates of differential fertility: a longitudinal study. *Amer. Psychologist*, 1955, **10**, 319. (Abstract)
68. MOSHINSKY, PEARL. The correlation between fertility and intelligence within social classes. *Sociol. Rev.*, 1939, **31**, 144–165.
69. NISBET, J. D. *Family environment: a direct effect of family size on intelligence*. London: Cassel, 1953.
70. NISBET, J. D. Family environment and intelligence. *Eugen. Rev.*, 1953, **45**, 31–40.
71. O'HANLON, GERARD S. A. An investigation into the relationship between fertility and intelligence. *Brit. J. educ. Psychol.*, 1940, **10**, 196–211.
72. PAPAVASSILIOU, I. T. Intelligence and family size. *Popul. Stud.*, 1954, **7**, 222–226.
73. PEARSON, K., & MOUL, MARGARET. The problem of alien immigration into Great Britain, illustrated by an examination of Russian and Polish Jewish children. *Ann. Eugen.*, 1925, **1**, 5–127.
74. PENROSE, L. S. The supposed threat of declining intelligence. *Amer. J. ment. Defic.*, 1948, **53**, 114–118.
75. PENROSE, L. S. Genetical influences on the intelligence level of the population. *Brit. J. Psychol.*, 1950, **40**, 128–136.
76. PENROSE, L. S. Propagation of the unfit. *Lancet*, 1950, **259**, 425–427.
77. ROBERTS, J. A. FRASER. Intelligence and family size. *Eugen. Rev.*, 1939, **30**, 237–247.
78. ROBERTS, J. A. FRASER. The negative association between intelligence and fertility. *Hum. Biol.*, 1941, **13**, 410–412.
79. ROBERTS, J. A. FRASER, *et al.* Studies on a child population. III. Intelli-

gence and family size. *Ann. Eugen.*, 1938, **8**, 178–215.

80. SCOTT, EILEEN M., & NISBET, J. D. Intelligence and family size in an adult sample. *Eugen. Rev.*, 1955, **46**, 233–235.
81. SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION. *The intelligence of Scottish children: a national survey of an age group.* London: Univer. of London Press, 1933.
82. SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION. *The trend of Scottish intelligence.* London: Univer. of London Press, 1949.
83. SCOTTISH COUNCIL FOR RESEARCH IN EDUCATION. *Social implications of the 1947 Scottish mental survey.* London: Univer. of London Press, 1953.
84. SEWELL, W. H. Differential fertility in completed Oklahoma farm families. *Amer. sociol. Rev.*, 1944, **9**, 427–434.
85. SHUEY, AUDREY M. Intelligence of college women as related to family size. *J. educ. Psychol.*, 1951, **42**, 215–222.
86. STÖCKLE, O. Die kinderreich Familie in ihrer sozialpädagogischen Bedeutung. *Heilpädag. Werkbl.*, 1954, **23**, 144–149.
87. SUTHERLAND, H. E. G. The relationship between IQ and size of family. *J. educ. Psychol.*, 1929, **20**, 81–90.
88. SUTHERLAND, H. E. G. The relationship between IQ and size of family in the case of fatherless children. *J. genet. Psychol.*, 1930, **38**, 161–170.
89. SUTHERLAND, H. E. G., & THOMSON, G. H. The correlation between intelligence and size of family. *Brit. J. Psychol.*, 1926, **17**, 81–92.
90. TABAH, L., & SUTTER, J. Le niveau intellectuel des enfants d'une même famille. *Ann. hum. Genet.*, 1954, **19**, Pt. 2, 120–150.
91. TERMAN, L. M., *et al. Genetic studies of genius*, Vol. I. Stanford, Calif.: Stanford Univer. Press, 1925.
92. THOMSON, G. H. The trend of national intelligence. *Eugen. Rev.*, 1946, **38**, 9–18.
93. THOMSON, G. H. Intelligence and fertility in Scotland. *Eugen. News*, 1950, **35**, 23–24.
94. THOMSON, G. H. Intelligence and fertility; the Scottish 1947 survey. *Eugen. Rev.*, 1950, **41**, 163–170.
95. TUDDENHAM, R. D. Soldier intelligence in World Wars I and II. *Amer. Psychologist*, 1948, **3**, 54–56.
96. UNITED NATIONS, DEPT. OF SOCIAL AFFAIRS. Population Division. *The determinants and consequences of population trends.* ST/SOA/Series A Popul. Stud. No. 17. New York: United Nations, 1953.
97. UNITED NATIONS. DEPT. OF SOCIAL AFFAIRS. Population Division. *Proceedings of the World Population Conference, Rome, 1954.* New York: United Nations, 1955.
98. UNITED STATES DEPT. OF COMMERCE, BUREAU OF THE CENSUS. Fertility: April, 1947. *Current Popul. Rep.: Popul. Characteristics*, 1948, Series P-20, No. 18, 1–21.
99. UNITED STATES DEPT. OF COMMERCE, BUREAU OF THE CENSUS. Fertility of the population: April, 1952. *Current Popul. Rep.: Popul. Characteristics*, 1953, Series P-20, No. 46, 1–22.
100. VERNON, P. E. The quality of the population. *Brit. J. educ. Psychol.*, 1950, **20**, 35–42.
101. VERNON, P. E. Use of intelligence test scores in population studies. *Eugen. Quart.*, 1954, **1**, 221–224.
102. WARBURTON, F. W. Relationship between intelligence and size of family. *Eugen. Rev.*, 1951, **43**, 36–37.
103. WARBURTON, F. W. Relationship between the intelligence of students and size of family. *Eugen. Rev.*, 1952, **43**, 188.
104. WESTOFF, C. F. The changing focus of differential fertility research: the social mobility hypothesis. *Milbank Mem. Fund Quart.*, 1953, **31**, 24–38.
105. WHEELER, L. R. A comparative study of the intelligence of East Tennessee mountain children. *J. educ. Psychol.*, 1942, **33**, 321–334.
106. WHELPTON, P. K., & KISER, C. V. Social and psychological factors affecting fertility; differential fertility among 41, 498 native-white couples in Indianapolis. *Milbank Mem. Fund Quart.*, 1943, **21**, 221–280.
107. WHELPTON, P. K., & KISER, C. V. (Eds.) *Social and psychological factors affecting fertility.* Vol. 1–4. New York: Milbank Memorial Fund, 1946–1955.
108. WILLOUGHBY, R. R. Fertility and parental intelligence. *Amer. J. Psychol.*, 1928, **40**, 671–672.
109. WILLOUGHBY, R. R., & COOGAN, MARGUERITE. The correlation between intelligence and fertility. *Hum. Biol.*, 1940, **12**, 114–119.
110. WILLOUGHBY, R. R., & GOODRIE, MIRANDI. Neglected factors in the differential birth rate problem. *J. genet. Psychol.*, 1927, **34**, 373–393.

Received June 6, 1955

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 3, 1956

STUDIES OF BRAIN DYSFUNCTION IN SCHIZOPHRENIA[1,2]

GLEN A. BRACKBILL
Veterans Administration, Palo Alto

In 1924, Charles Dunlap (**24**) stated quite unequivocally that there were no organic lesions or histopathological findings which could be shown to be characteristic of schizophrenia. According to him the changes which had been reported were not essentially different from nerve cell changes found in nonpsychotic persons. This rather sweeping conclusion was based on his comparison of the brains of eight schizophrenic patients with those of four "control" subjects. His control group consisted of a bootlegger who had been shot through the lung and died shortly thereafter, two men who died of arsenic poisoning, and one woman who died of peritonitis. It appears, contrary to Dunlap's statement, that there is room to question the "normality" of such a group. At any rate, this pronouncement apparently had an inhibiting effect on further investigations into the role of organic factors in schizophrenia. Few such investigations were undertaken in the decade following Dunlap's authoritative statement. In 1930, Freeman (**33**) quoted Dunlap's conclusions but was willing to report finding a deficiency of catalytic iron in the nerve cells of schizophrenics. Here again seems to be evidence of the profound effect of a statement backed by an authority figure which blocked progress in an area deserving of continuing experimental study.

Perhaps the electroencephalographers were unimpressed by authority or were not aware of it. At any rate, they appear to have been, in large part, responsible for rearousing interest in the search for an organic factor in schizophrenia. In the process of recording the brain waves of various groups of people, they found among schizophrenics waves resembling those found in cases of known brain damage. Histopathologists and biochemists have followed up these leads so that the recent psychiatric literature presents a number of provocative studies. It is the purpose of this review to present some of these recent findings and to suggest some hypotheses.

While it is probably true that most of the work in the physiology, genetics, and psychology of psychoses is at least indirectly related to the subject of brain dysfunction, the electroencephalographic, histopathological, and biochemical studies are both directly relevant and of sufficient importance in themselves to merit concentrated attention. The review will include these studies as well as several psychological and physiological studies which focus specifically on the question of brain disorder in schizophrenia. Studies made within the last twenty to twenty-five years will be included.

It is felt that psychologists could make a worth-while contribution to the studies of possible central nervous system pathology among schizophrenics. Measurement of organic damage by various kinds of psychological tests has been rewarding to the extent that tests have been shown to be useful in research and in theoreti-

[1] From the Veterans Administration Hospital, Palo Alto, California.

[2] I would like to thank Mary Ann Swanson and Elizabeth Grisak for their helpful assistance.

cal formulations about the nature of thinking in cases of cerebral damage. In many clinics and hospitals neurologists have come to depend on psychological tests in the diagnostic study of central nervous system involvement. It is hoped that this review will stimulate research interest on the part of psychologists who are concerned with the physiological correlates of behavior.

EEG Studies

In 1937, Travis and Malamud (**76**) reported a study unfavorable to the hypothesis that schizophrenics show patterns different from the normal. These investigators were apparently only secondarily interested in this possibility since they were more interested in testing some of Berger's original formulations about the nature of the EEG. However, they did find no differences in either amplitude or frequency of the waves of nine schizophrenics compared with those of a group of seventy-five normals and a group of stutterers. They were aware of the small size of their group of schizophrenics, and are tentative in their conclusions.

In the following year, Gibbs, Gibbs, and Lennox (**34**) found that epilepsy is present about three times as often among schizophrenics as in the general population. Gibbs *et al.* point out that their experience shows that the majority of schizophrenics they have tested show abnormalities of rhythm. They found both fast and slow EEG patterns. Although they do not present statistics, they were impressed by the similarity of many of the records of schizophrenics to those of psychomotor epilepsy. This finding should arouse some interest in the possibility of an organic factor in those cases in which an epileptic EEG is found. The meaningfulness of such a question would, however, depend upon the acceptance of an organic basis for psychomotor seizures.

Jasper, Fitzpatrick, and Solomon (**48**) studied the EEG's of 82 schizophrenics, so diagnosed by agreement of five psychiatrists (no small accomplishment in itself). They also included a group of 51 epileptics and 60 normal subjects. They found, as is expected of schizophrenic records by now, no specific EEG pattern, but rather, a wide range of variability. Contrary to the study of Travis and Malamud, they found a variety of anomalies to a more extreme degree and in a greater number among schizophrenics than normal subjects. Specifically, they report that 23 per cent of the cases diagnosed as schizophrenic showed either clinical or electroencephalographic evidence of brain activity similar to that characterizing the epileptic. They are somewhat hesitant in their conclusions about the role of organic factors in producing the abnormal EEG's. They state, "15% of our group of patients diagnosed as schizophrenic may well be suffering from mental disorders due to head injury, mental deficiency, or other organic defects of the brain" (**48**, p. 849).

Another study involving 52 chronic schizophrenics and 500 control cases reports essentially the same results. Davis and Davis (**19**) also feel that schizophrenic patients do not have a typical "schizophrenic EEG" in the same sense as, for example, the spike and slow wave of petit mal epilepsy. They found variations in wave patterns outside the range of normal subjects. They come to much the same conclusions as Jasper *et al.*, but are more cautious in their concluding statement. "The quality of the atypical variations resembles that of the changes which occur in the electroencephalograms of persons known

to have brain lesions or who suffer from the various forms of epilepsy, or normal persons during sleep, or when breathing an inadequate supply of oxygen" (19, p. 1019). They permit the reader to draw his own conclusions about the meaning of their findings in regard to possible organic factors in some schizophrenics.

In 1940, Rubin (**70**) published EEG studies of 14 schizophrenics, 2 manics, and 1 case of traumatic psychosis. He used the per cent time alpha as a measure. By studying the distribution of per cent time alpha over the various areas of the cortex, he felt that he was able to detect the presence of atrophy in these cases. His diagnosis was confirmed in 10 of the cases by pneumoencephalograms. Air studies had not been done on the remaining subjects. Rubin states not only that there was central nervous system pathology present, but also that the nature of this pathology was atrophy. Rubin's results were confirmed by Moore, Nathan, Elliott, and Laubach (**55**). In the majority of their sixty cases they found evidence of atrophy, enlargement of the ventricular system, and cisternae. A study of the films suggested that there was a selective atrophy involving the parietal lobe and insula. Unfortunately they do not report on the frequency of occurrence. It could be inferred that at least 31 of the cases showed such atrophy. Again it is difficult to assess these findings since the authors used no control group and hence could not evaluate the frequency of atrophy among nonschizophrenics.

Davis (**20**) studied the EEG's of 132 schizophrenics in an attempt to work out a classification system for schizophrenic EEG's. She found that she could isolate three general types of patterns. Group I showed essentially no differences from the normal. Group II showed dysrhythmias without convulsive characteristics. Group III contained records which suggested the possibility of localized or generalized irritative lesions, cysts, atrophy, or other pathological conditions. This study again suggests that an appreciable number of schizophrenics show electrical activity of the cerebral cortex which has long been associated with some form of structural brain damage.

Finley and Campbell (**31**) studied the EEG's of 500 schizophrenics and 250 normals using a triple classification similar to that of the previously mentioned investigator. They categorized the records as normal, borderline, or abnormal. When these classifications were applied to their two groups they found 40 per cent of the schizophrenics had normal records while 70 per cent of the normal group had normal records. Twenty-eight per cent of the schizophrenics had abnormal records while 7 per cent of the normals had abnormal records. The remaining subjects had records in the borderline category. Thus the incidence of abnormal tracings among schizophrenics was twice that of the normal group. Finley and Campbell found that 30 per cent of the patients' records showed tracings indicative of a disorder of function in the central nervous system. They point out, however, that the EEG pattern is not diagnostic of schizophrenics since the kinds of wave activity found are not unique to this group.

Kennard and Levy (**50**) found among 100 schizophrenic patients exactly the same 60 per cent of abnormal patterns as reported by Finley and Campbell. They looked for correlations between the abnormal patterns and the patients' clinical pictures. The incidence of abnormal EEG's was higher for those patients

with early onset and prolonged illness. If the illness had lasted three years or less, the incidence of abnormal tracings was only 42 per cent as compared to 72 per cent if the illness had persisted over ten years. This relationship also seemed to hold in regard to the degree of intellectual impairment shown by the Wechsler-Bellevue Test. It is suggested that these findings are consistent with the idea that " . . . the schizophrenic process [can be] thought of as a progressive disorder which ultimately profoundly affects the performance of all organic systems including that of the cerebral cortex" (**50**, p. 421).

Grinker and Serota (**35**) implicate the hypothalamus in their EEG study of schizophrenia. Their report is in the nature of an hypothesis resulting from the observation of some autonomic reaction in schizophrenia. There was little electrical activity in the hypothalamus and cortex when their patients were subjected to external cold. They also found that electrical stimulation of the hypothalamus failed to arouse activity in that structure or in the cortex. Among neurotics and normals an increase in activity was apparent. They suggest that these findings indicate a physiological failure at the cortical level because the usual hypothalamic driving force is lacking. The inference could be made from these formulations that the thalamocortical circuits are not functioning so that the cortex is deprived of impulses from lower centers. The Tulane group headed by Heath (**38**) has also suggested that subcortical centers may be implicated in schizophrenia. They report spiking activity from areas near the thalamus in the septal region. EEG's from leads implanted by the Horsley-Clarke apparatus in subcortical areas of schizophrenic patients often showed the spiking activity generally believed indicative of structural damage. They apparently felt that if they succeeded in eliminating the spike by administering electric shock through the subcortical electrodes, the patient had a better chance for recovery than if the spike activity continued.

One other EEG study is of incidental interest. Turner and Lowinger (**77**) saw a marked difference in prognosis for recovery of schizophrenics given shock treatment, depending on the preshock EEG pattern. None of the patients with abnormal preshock waves was able to leave the hospital after treatment. One-fourth of those with normal patterns were discharged following the treatment. Again, a relationship between the presence of abnormal wave activity and malignancy is suggested.

In summary, it is interesting to note the progressive changes in the significance ascribed to EEG findings in the research on schizophrenia. The first studies emphasize lack of consistency in the tracings. In the later studies, greater confidence is placed in the idea that EEG studies have demonstrated brain dysfunction in an appreciable number of cases of schizophrenia, despite variability in the tracings. Finally, some investigators have been willing to state the nature of the difficulty, i.e., atrophy of the cortex and/or failure of thalamocortical circuits. The histopathologists have been much more specific in their statements regarding the nature of the brain pathology found. This larger group of studies will be reviewed next.

Histopathological Studies

One of the persistent proponents of organic factors in schizophrenia is Eugen Bleuler (**12**). His monograph on the schizophrenias is no doubt still the outstanding text in the field. It

was his early clinical observations of sensory disorders among such patients that led him to suggest organic dysfunction of the brain. In 1930, Bleuler wrote a provocative essay (**13**) in which he tried to sort the psychogenic from the physiogenic in schizophrenia. It seems worthwhile to quote part of his conclusions in regard to the role of the central nervous system in schizophrenia:

> Many cases of stupor, with their general prostration of the elementary psychic functions, conception and train of thought, often point clearly to brain pressure and, on autopsy tense edema of the pia or brain swelling is found. In the various forms of such deliria, the fundamental similarity of certain symptoms or of the whole picture to other physiogenic conditions, intoxication, fever psychoses, epileptic absences, meningitis, encephalitis, cannot be denied, and in all such cases we also find in the autopsy histological alterations of the brain tissue which show some uniformity. But in all chronic cases too, decreases in the amount of ganglion cells and certain changes in the glia, furnish proof that we are in the presence of a brain lesion . . ." (**13**, p. 206).

There are serious difficulties in experimental histopathology, many of which have not been overcome. For instance, as Spielmeyer (**74**) points out, it must be remembered in using necropsy material that the diseases which cause death may themselves bring about changes in the brain. Also, it is not known what the agonal process itself may do to brain cells. Neither of these variables has been thoroughly explored. It might be added that although biopsy material can now be obtained easily at the time of prefrontal lobotomy, control material from normal subjects is not so readily procured. If control material is not obtained, it is impossible to assess the extent to which the results may be attributed to the experimental procedure. Many of the studies to be reviewed in this section have not overcome these difficulties altogether. Hence, the degree of confidence to be placed in the conclusions of these studies is necessarily limited, as is also the generalizability of the results.

In 1930, Spielmeyer (**74**) recognized these potential sources of error, but still felt justified in concluding that characteristic histological changes were evident in schizophrenia. The finding he emphasizes is the loss of cells in the cortex. He also reports changes in glial cells and the presence of destruction products. He concludes that there is sufficient evidence to state definitely that schizophrenia is an organic process.

In 1932, Bamford and Bean (**6**) studied a series of cases they describe as acute dementia praecox. These authors report on three cases but base their conclusions on an unspecified number of additional cases. Their conclusion is quite different from Spielmeyer's. In fact, this study is one of the few since 1930 reporting completely negative results. (It should be remembered, of course, that negative results are less likely to be reported than are positive findings.)

As is often found in the psychiatric literature, there are a number of case studies reported. Several of these will be reviewed even though they do not represent well-controlled nor truly experimental studies. It is perhaps not necessary to state that fruitful hypotheses may be derived from observation of the single case.

Ferraro (**29**) reports two cases diagnosed dementia praecox. He found swelling of the oligodendroglia in both cortical and subcortical tissue, but enlargement was most prominent in frontal and temporal areas. He also noted demyelinization, perivascular infiltration, and the presence of gitter cells. These were taken to be indications of a degenerative process not to be expected in the normal

brain. Ebaugh, Barnacle, and Neuberger (25) detected histological alterations in fatalities resulting from electroshock therapy. They found changes in astrocytes and in the neuroglia, and suggest that although these modifications may have been related to the shock they were not limited to areas of strongest current. Their conclusions are not explicitly stated, but they imply that shock alone could not account for the brain damage. Ferraro (30) reports another case which had been diagnosed as schizophrenia for more than ten years. The clinical picture was clearly that of a schizophrenic psychosis, but on autopsy a diffuse demyelination of the brain was found. The disease was apparently so widespread and progressive that it could scarcely be attributed to a sudden illness. Polatin, Eisenstein, and Barrera (66) describe two typical cases of chronic schizophrenia, similar to those of Ferraro. Their discovery of slow wave activity in the EEG's was the first clue to the presence of organic damage; pneumoencephalography revealed cerebral atrophy; and the Rorschach suggested organic brain disease. Biopsy studies revealed what the authors considered to be irreversible changes in the cerebral cortex.

Roizan, Moriarity, and Weil (68) examine the case of a 34-year-old woman who died after a brief acute psychosis diagnosed as catatonic schizophrenia. Here, as in Ferraro's research, study of the brain tissue revealed a demyelinating process in the central nervous system. These authors interpret their findings not as indicating that schizophrenia is an organic process but that impaired brain function acts as a precipitating factor for a schizophrenic reaction.

Holt and Tedeschi (45) and Van der Horst (78) report—as did the previous investigators—the presence in individual cases of demyelinization, microglial and oligodendroglial hyperplasia, loss of cells, and vascular infiltration. Additionally, Holt and Tedeschi detected cytoplasmic involvement in the nerve cells themselves along with nuclear disturbances and lipoid deposits.

One of the most interesting and best controlled studies was made in 1944 by Kirschbaum and Heilbrunn (51). They took biopsies from the frontal lobes of eleven chronic schizophrenics and also from three normal subjects. Biopsies from several cats and rats provided further control material. For the experimental material they report ganglion cell degeneration as well as alterations in glial cells and blood vessels. These changes are similar to those found in chronic intoxication and metabolic disorders. Comparison of experimental and control material indicated that the results could not be attributed to the experimental procedures. This study seems most impressive because many of the methodological flaws of other studies were avoided: The tissue was taken from the living organism and studied immediately, thus ruling out the complicating effects of the agonal process as well as the disease which might cause death. Also, the use of controls in this study held constant any effects the anesthetic and surgical procedures might have had.

Another well-executed biopsy study is reported by Elvidge and Reed (27). They used material from 19 psychotic patients, 13 of whom were schizophrenic. Biopsies were also obtained from 16 control subjects, mostly epileptics, and from 14 laboratory animals. They too report for the schizophrenic subjects changes in glial cells, with hypertrophy and swelling. In some cases the oligoden-

droglia showed nuclear modification. These changes were most apparent in the white matter; in some cases they were generalized, while in other cases they appeared more patchy. Similar changes were found for those control cases in which the patient showed clouding of consciousness between epileptic seizures, but not for those cases in which the patient remained clear and oriented between seizures. A number of the psychotics were followed over periods of from one to two years, and the biopsies were repeated. The findings were the same as before. It is unfortunate that none of the cases had remitted between biopsies, for if it could have been demonstrated that the changes had slowed down or ceased in recovered cases, it would have been an impressive argument for the importance of organic factors in schizophrenia. Elvidge and Reed suggest that their results indicate disturbances or interruptions of commissural pathways in the brain, and that it is this "disruption" of impulses to and from various parts of the brain which causes disturbances in intellectual and emotional reactions.

Winkelman and Book (**80**) made a post-mortem study of 10 cases of "typical" schizophrenia. Their findings are listed as (*a*) focal cell loss, (*b*) a general decrease in the number of nerve cells, especially in the anterior half of the brain, (*c*) chronic cell diseases, including cell shrinkage and ghost cells, (*d*) amount of fat in cells usually greater than expected, (*e*) increase in astroglia, and (*f*) mild general demyelinization in subcortex. They claim that the changes found in the brains of their patients were much more severe quantitatively than those usually found in nonpsychotic cases.

From another post-mortem study come similar results. Rupp and Wilson (**71**) studied 37 patients, all under fifty years of age. Biopsies revealed loss of cells, gliosis, atrophy, edema, vascular lesions, and areas of softening. The authors do not report data for control subjects.

Probably the most persistent investigators in the histopathology of schizophrenia are Papez, and Papez and Bateman (**57**, **58**, **59**, **60**, **61**, **62**, **63**). Their methods and techniques are carefully detailed and appear to meet the standards of good research, with the exception that the extent to which they have studied control material is not always made explicit. In 1944, Papez (**57**) reported finding dark staining particles, inclusion bodies, in the cytoplasm of nerve cells. According to him, these particles were not artifacts produced by the technique but were living microorganisms of a parasitic nature. In 1949, Papez and Bateman (**59**) made another study of biopsies. In their sample of 50 patients, there were 42 schizophrenics, 6 manic depressives, and 2 syphilitics. They identified three stages of cell disease. In the first stage, the cells showed intact cytoplasm occupied by a small number of the inclusion bodies. Cell nuclei showed some deformity. The second stage revealed highly vesicular nuclei, enlarged nucleoli, and cytoplasm filled with inclusion bodies. The third stage was characterized by naked nuclei stripped of cytoplasm. The inclusion bodies had apparently invaded the walls of small blood vessels, producing degenerative changes. The authors studied these inclusion bodies under a dark phase microscope and concluded that they were dealing with a living organism. They were able to watch these zooid organisms in various phases of the life cycle and to keep them alive over a period of many hours in a suspension of cortical material. Drawings made from mi-

croscope slides are presented in the article. They indeed suggest an ominous picture of parasitic infestation. Papez and Bateman propose and argue convincingly for the thesis that the histopathological changes reported by other investigators are associated with various phases in the life cycle of the organism.

Schaderwald (**72**) confirmed many of the findings of the previous investigators, using material from 34 biopsies taken at the time of prefrontal lobotomy. He states that his data is in complete agreement with that reported by Papez, and Papez and Bateman. However, because of the lack of controls, he is much more cautious about ascribing pathognomonic significance to the findings.

The reviewer must agree with Weinstein (**79**) in discussing the studies of Papez, and Papez and Bateman: "Merely the unusual nature of the claims must not be allowed to interfere with open-minded considerations of the findings . . . . Many questions remain which can be answered only through carefully designed experimental work with adequate attention to controlled observations . . . " (**76**, p. 549).

The thalamus and autonomic nervous system have also been implicated as sites of possible lesions in schizophrenia. In 1939, Stein and Ziegler (**75**) reported a biometric analysis of 24 patients and 1 normal subject. The brains and thalami in the schizophrenic patients were smaller and the cell count lower than for manic-depressive patients in general. However, none of the differences was statistically significant. The investigators felt that the small size of their sample contributed to the negative findings.

For a group of 16 paranoid schizophrenics, Bateman and Papez (**7**) found inclusion bodies in thalamic cells as well as cell loss as high as 90 per cent in the association nuclei. On this finding they based their proposition that hallucinations and delusions are the result of abnormal nervous discharge of the thalamus to the cortex. Since sensory input from peripheral sense organs must pass through the thalamus, the presence of diseased cells there could distort the impulses to such an extent that the misperceptions characteristic of psychoses would be expected.

Two reports question the importance of histopathological findings in schizophrenia. In 1949, Rowland and Mettler (**69**) studied 22 schizophrenics and 1 manic patient. They hypothesized that if cell loss is actually an important feature in psychosis, it should be positively correlated with severity of illness. However, their findings indicated no difference in cell count between a group hospitalized more than twenty-two months and a group hospitalized less than twenty-two months. They contend that the various methods of counting cells are frequently of questionable reliability, and suggest that this fact in large measure accounts for the conflicting reports of cell loss. The same opinion is voiced by Wolf and Cowan (**81**). They add that perhaps the positive results which have been reported can be explained on the basis of misjudgment of the limits of normal variation in brain tissue, and on the basis of methodological artifacts. They feel that available studies are not sufficient to permit accepting the thesis of a specific organic factor in schizophrenia. This position is vigorously opposed by Winkleman in a discussion following their article.

In summary, it can be stated that a number of different pathologists report similar results. A frequent finding is loss of nerve cells. Pneumo-

encephalographic reports of atrophy appear to support this impression. Other repeated findings are: the presence of a demyelinizing process, degenerating nerve cells, and pathological changes in the various glial cells. It seems fairly well established that the brains of some schizophrenics show structural damage, but the significance of the damage is difficult to evaluate because comparable studies of nonpsychotic subjects are generally lacking.

Biochemical Studies

So far, very few biochemical studies of brain function in psychosis have been made. In 1930, Freeman (**33**), after agreeing with Dunlap that there were no histopathological findings in schizophrenia, reported a deficiency of catalytic iron in the brain cells of schizophrenics. His sample included 35 schizophrenics and 16 patients of other diagnoses. Freeman suggests that the lack of this catalytic agent (which is necessary for the use of oxygen) may underlie some of the symptoms. An additional inference that could be made from these findings is that the inability of the cells to take up oxygen would eventuate in their degeneration.

Doust (**23**) used two samples of schizophrenics, one English and one American, totaling 87 patients. He studied the brains by means of the spectroscope and photoelectric oximetry, and found marked reduction in the effective capillary oxyhemoglobin. In view of the extreme sensitivity of nervous tissue to anoxia and the irreversibility of the consequent damage, Doust's results seem quite important to the question of central nervous system involvement in schizophrenia. Also, it is interesting to note that this investigator reports a much more profound deficiency in capillary oxyhemoglobin among the so-called "process" schizophrenics (i.e., classical dementia praecox) than for any other group studied. The reviewer feels that this finding has considerable significance for methods of selecting groups of schizophrenics for study. This point will be elaborated in the discussion to follow.

Physiological Studies

From England, Hoffer, Osmund, and Smythies (**44**) report the accidental discovery of adrenochrome, or pink adrenaline. This substance, closely related to adrenaline, is a deterioration product of that hormone. Although chemically it is only slightly different from adrenaline, its effect on the body is quite different. It shows toxic effects in much the same way as does mescaline. For this reason, adrenochrome caught the interest of investigators hypothesizing toxicity as a basis for schizophrenia. (The catatonic subgroup has usually been the object of such study, since they present many of the symptoms generally indicative of toxic disorders.)

Hoffer, Osmund, and Smythies studied the effects of pink adrenaline on a group of volunteer normal subjects. The results are striking. When given adrenochrome, their subjects showed behavioral symptoms of schizophrenia. Furthermore, their subjective reports confirmed the observer's impressions that they were delusional and experienced hallucinations. EEG's taken on the subjects under the influence of the drug showed arrhythmias and epileptic patterns within half an hour from the time of administration. The authors hypothesize that when adrenochrome enters the cerebral cells it inhibits carbohydrate metabolism and interferes with cell respiration. In schizophrenics, an excess of adrenochrome may be produced, resulting

in autointoxication. One might also speculate about a possible connection between an excessive production of adrenochrome and the hypothesis that schizophrenics suffer a chronic hyperactivity of the sympathetico-adrenal system.

The endocrine system has been considered the site of possible pathology in schizophrenia for well over half a century. Every gland from sex to pituitary has at one time or another by various investigators been identified as the locus of the difficulty. The findings have been tantalizing but equivocal, and the work done in this area would constitute a bibliography several times larger than the present one. However, a few of the studies which explicitly link endocrine and autonomic nervous system function will be mentioned.

Hoskins (**46**), in reviewing the literature on endocrine function in schizophrenics, reports some of the work done over a period of twenty years at Worcester. His observations led to the conclusion that in an appreciable number of cases thyroid dysfunction is a basic factor. Some schizophrenic patients seem to suffer from a lack of thyroid secretion; however, continued treatment with the hormone raised the question of whether the target tissue was not at fault. That is, since some patients did not respond even to massive doses of thyroxin, Hoskins concluded that the difficulty is not so much in thyroid gland production itself as it is in an apparent failure of the appropriate receiving or "target" tissue to utilize the hormone.

The other glands which frequently show anomalous reactions are the adrenals. The function of the adrenals is relevant to our subject, for while there is no direct nervous control of the adrenal cortex, its function is intimately related to that of the autonomic nervous system. Indirectly, the secretions of the adrenal cortex are influenced by the autonomic nervous system through the hypophysis, which is controlled by the hypothalamus. In the past few years there has appeared an impressive number of studies on adrenal function in schizophrenia. No doubt the research on adrenal activity has been accelerated because ACTH and cortisone have become available as research tools.

Hoskins (**46**) suggests that in schizophrenia the adrenal-autonomic mechanism is somehow disturbed, probably ultimately expressing itself in altered metabolism. Work on the adrenals and the emergency functions has been followed up by Hoagland and others (**17**, **39**, **40**, **41**, **42**). These studies are typical of the many reported. The primary finding appears to be that the schizophrenic exhibits a physiologically subnormal response to stress. Altschule, Promesel, and Parkhurst (**3**) found that the reactions of schizophrenics to ACTH injections are not reduced, so they propose that the difficulty lies in the hypothalamus. Perhaps this vital center fails to activate the pituitary to secrete sufficient adrenocorticotropic hormone, resulting in a reduction of the usual adrenal stress reaction. Hoagland *et al.* (**39**) support this point of view.

After a rather thorough review of the literature on the bodily functions in psychoses, Altschule (**2**) concludes that most of the disordered physiology reported is a consequence of the psychosis rather than an etiological factor. He makes one exception, namely that disturbances in adrenocortical function may cause metabolic disturbances which interfere with brain function, producing psychoses. He emphasizes that there has been a

general tendency to ignore the physiology of mental illness. This tendency has been reinforced by the continued use of the term "functional," which implies, wrongly, a lack of physical or biological imbalance. Actually, there is no such thing as a *functional* psychosis. Altschule states in regard to this point: "Theories that ignore cerebral metabolic processes and regard psychosis as merely a state of mind, deliberately chosen by the patient because of environmental influences, have nothing—not even reasonableness—to support them" (**2**, p. 212).

Paralleling the endocrine studies is a series of experiments with various drugs which produce "model-psychoses" (**32**). Studies by DeJong (**21**) with mescaline, bulbocapnine, and other catalepsy-producing drugs led him to hypothesize a "faulty detoxification theory" of catatonia. According to him, the liver is the organ at fault.

Since DeJong's studies, numerous other drugs have been used in an attempt to find a toxic agent responsible for psychosis. Of these, one which has recently received considerable attention is lysergic acid diethylamide (LSD-25) (**22**, **32**, **43**, **47**, **52**). This drug seems to act on the hypothalamus and in normals, produces many of the psychological effects characteristic of the schizophrenic psychoses. The drug has also been administered to schizophrenics, and the reported effects vary. In general, however, the effect seems to be one of exacerbating existing symptoms. For normals, these effects are reported: visual disturbances, thought disorders, mood changes, paranoid tendencies, auditory phenomena, etc. Physiological changes are also noted. These changes are principally those associated with autonomic activity. This leads to the hypothesis that it is the homeostatic balance which is disburbed by the drug (**67**).

On the negative side, Sloane and Doust (**73**) interpret their results as indicating that schizophrenics show a diminished or unchanged autonomic reactivity following administration of LSD-25. Their sample included 19 patients and 14 controls. In an attempt to clarify the specific action of the drug, Mayer-Gross, McAdam, and Walker (**52**) found that it produced changes in phosphate metabolism. They propose that the apparent increased respiration of brain tissue may account for the psychological disturbances produced by the drug.

It is important to note that the last studies reviewed are also of methodological interest, for when a disease can be artifically produced, it is an important step in the direction of control of that disease. The chemist can understand his compound completely, not when he has succeeded in analyzing it but only after he can synthesize it in his laboratory. He is then in a position to control it and predict its action. It is evident, of course, that the reactions produced by the administration of a drug to normal persons cannot at this stage be unequivocally equated with schizophrenic psychoses. For that matter, as is pointed out by Pennes (**64**), the reactions to LSD-25 cannot be entirely due to the *specific* action of the drug; otherwise, the wide variability in its effects would not have been found. He too found that not all of his 55 schizophrenic subjects reacted to the drug in the same way. Actually, in some cases the drug seemed to have a temporary normalizing effect. The most judicious conclusion is probably that response to the drug is the result of interaction between psychological and physiological factors.

In spite of the large amount of work done with various drugs, the mechanism of their action is still not well understood. Nevertheless, more and more of them are being discovered and tried out. Some, as reported by the press, have assumed the status of wonder drugs. Reserpine (**54**) and chlorpromazine are two such. Again, the effectiveness of these drugs seems primarily ascribed to their action on the autonomic nervous system (**36**).

Psychological Studies

It is rather disconcerting to find that schizophrenic patients often respond to psychological tests in ways that we have learned to consider characteristic of patients with known organic involvement. This presents a difficult problem to the clinician who is asked to make a differential diagnosis. In the past, test constructors have been busy trying to show how well their tests differentiate the current nosological groups. Little attention has been paid to the degree of overlap, even though the similarities between two groups on test performance may be actually more impressive than the differences. One exception is the report of Hanfmann and Kasanin (**37**) in the discussion of their widely known test of conceptual thinking. They found a small but significant mean difference between groups of known organics and schizophrenics. Regarding the response similarity of schizophrenics to organics, they state, "Cases of this type often bear a close resemblance to cases of irreversible brain disease, except for a greater variability and a certain imaginative quality of their productions which are lacking in organic cases" (**37**, p. 75). It might be added that this imaginative quality is often very difficult to detect.

Altrocchi and Rosenberg (**1**) also found overlap of schizophrenics and organics in their work with a new block test of conceptual thinking. Some of the so-called functional psychotics could not be differentiated from known organic cases, on the basis of the test, although others could. The hypothesis is immediately suggested that among the heterogeneous group of mental ills called "schizophrenia," there is a considerable percentage suffering chronic brain damage.

In an attempt to test this hypothesis, Brackbill and Fine (**14**) studied three groups of patients: a group of cases of known organic damage; a group of cases of classical or process schizophrenia, selected on the basis suggested by Kantor, Wallner, and Winder (**49**); and a group of reactive or acute cases. Using Piotrowski's (**65**) Rorschach signs of organic involvement, they found that the organic and process groups could not be distinguished from each other, while the reactive group showed significantly fewer organic signs than did either of the other two groups of patients. This study provides additional support for the thesis that among schizophrenics there is an appreciable number who suffer central nervous system pathology.

Another approach illustrative of the type of contribution which could be made by psychologists is that of Meadow and Funkenstein (**53**). They classified a group of 58 schizophrenic patients into three subgroups on the basis of autonomic reactivity. Group A had an autonomic pattern indicative of some kind of "release" from cortical inhibition, suggesting a breakdown in thalamo-cortical or cortico-thalamic interaction. On tests of abstract thinking these patients showed the severe impairment usually associated with brain damage. The general behavior of these

patients was of the nature which typically indicates poor prognosis. On the other extreme were the symptomatically schizophrenic patients of group B, who showed no disturbance in autonomic reactivity and no disturbance in abstract thinking, little anxiety, and no organized delusions. Prognosis was guarded, but was better than for group A. The C group showed severe anxiety, a different pattern of autonomic balance, and the best prognosis. It could be hypothesized that group A represented the classical or typical schizophrenic with organic symptoms, while the other two groups represented reactive or situational disturbances.

Hypotheses and Suggestions

The various investigators in this area have suggested a number of ideas and concepts in regard to the meaningfulness of these findings. Eickhoff (**26**) and Anderson (**4**) propose an organic factor in childhood schizophrenia. The former summarizes her observations as follows:

> I am therefore, postulating that schizophrenia in childhood is an arrest in the development of abstract thought and emotional maturity at an infant or toddler level; that this arrest is dependent basically upon a defect in the acquisition of general sensation; and this is due either to a defect in the neurological systems concerned with pain, touch, temperature, position, and vibration sense or to faulty stimulation from the outside or both; and this defect leads to a delay in the formation of the body image and other images" (**26**, p. 234).

Much the same statement is made by Anderson as the result of her observation of schizophrenic children:

> This outstanding modification in my conceptual thinking concerning schizophrenia would be in seeing the development of these reactions as being determined not by a rather ambiguous rejection behavior on the part of the mother, or any other significant person, but as a failure in interpersonal relations brought about primarily by a very specific type of organic brain deficit in the child. The exact nature of the physiologic pathology would probably be in the associational pathways of the most superficial layers of the cortex. The deficit might be of any degree from minimal to extensive (**4**, p. 37).

Lauretta Bender's (**10, 11**) long years of work with schizophrenic children has led her to a biological conception of schizophrenia. At times she speaks of it as an organic brain disease, proposing that there is an inherited weakness in the homeostatic mechanism. This weakness in autonomic function is such that the usual crises of physiological development may be sufficient to precipitate the illness. Follow-up studies of schizophrenic children showed that the symptoms closest to the biological were the best predictors of adult schizophrenia. This again bolsters the concept of a process schizophrenia, which starts early, is malignant, and is nonremitting.

All three of these investigators see evidence of brain damage even in these most early cases of schizophrenia. It is interesting that the defect is suspected to be somewhere on the input side of the nervous system, for Bateman and Papez (**7**) also suggest a distortion on the input side, resulting from cell disease in the thalamus. Also, according to Nielsen (**56**), there is a kind of schizophrenia that is based on a diencephalic lesion of a specific nature. He believes that the disease may eventually reach the cortex late in the process, and proposes that the defect is probably inborn, producing a disorganization of neuronal patterns. He proposes to call this an "apsychotic schizophrenia." Patients of this group are also called "ambulatory schizophrenics." They are nomadic, make a borderline schizoid adjustment, and their contacts with hospital and clinic are unproductive of any change in the pattern of adjustment.

Bychowski (**15**, **16**) speaks of schizophrenic thinking as similar to the kind of organic reactions described by Goldstein. After pointing out the concrete character and the similarity of thought processes for the two groups, he arrives at the hypothesis that there is a dynamic deficiency of cortical and subcortical organization in schizophrenia. He proposes that a certain group of schizophrenics suffer an agnosia and apraxia of thought based on cerebral deficit.

Bellak (**8**, **9**), after reviewing some 3,500 papers on schizophrenia, has formulated what he calls a "multiple factor theory." He maintains that the illness can be the result of a multiplicity of etiological factors. In one case it might be the result of liver pathology; in another, organic brain disease; etc. The symptomatology of schizophrenia may be a sort of "final common pathway" for the expression of illnesses produced by many different causes. This formulation is apparently an attempt to encompass the divergent—and at times confusing—findings in schizophrenia. In such a system, one could think of the behavioral symptomatology as the phenotype which is the expression of a number of different genotypical factors.

It is the reviewer's impression that the findings reported by these investigators merit more consideration from psychologists than they have so far received. It is probably true that the resolution of the question of an organic basis for schizophrenia rests with neuropathologists. There is, however, room for psychologists to make a contribution. It is widely accepted that the group now labeled "schizophrenic" is a heterogeneous one, often with little similarity among its members. One of the reasons for such conflicting research findings could well be this heterogeneity. That is, it might be hypothesized that some of the variability noted results from using samples of patients actually representing different populations. Therefore, it appears that one of the first steps in clarifying the problems of research in schizophrenia is to work out a more effective classification scheme than the four subgroup system now used. It is the reviewer's impression that the evidence is sufficient to postulate the existence of organic brain disease in some, but not all, schizophrenic patients. Those who show the classical signs of formal thinking disorder, such as concreteness, disconnected associations, etc., may have some kind of structural brain damage. On the other hand, those who show delusions, hallucinations, etc., but no thought disorder, may not have central nervous system pathology. Perhaps it is too early to make a definite statement regarding the exact nature and mechanisms of brain dysfunction, either etiologically or dynamically, but certainly there are a number of available theories of brain function which could be applied to test current hypotheses and to suggest further ones.

BIBLIOGRAPHY

1. ALTROCCHI, J., & ROSENBERG, B. G. An investigation of a new sorting technique for evaluating abstract behavior. (In progress.)
2. ALTSCHULE, M. *Bodily physiology in mental and emotional disorders.* New York: Grune & Stratton, 1953.
3. ALTSCHULE, M., PROMISEL, E., PARKHURST, B. H., & GRUENBAUM, H. Effects of ACTH in patients with mental disease. *Arch. neurol. Psychiat.* 1950, **64**, 641–649.
4. ANDERSON, CAMILLA. Organic factors predisposing to schizophrenia. *Nerv.*

Child., 1952, **10**, 36–42.

5. ANDERSON, E. W., & RAWNSLEY, K. Clinical studies of lysergic acid dilehylamide. *Monatsschr. Psychiat. U. Neurol.*, 1954, **128**, 38–55.
6. BAMFORD, C. B., & BEAN, H. A histological study of a series of cases of acute dementia praecox. *J. ment. Sci.*, 1932, **78**, 353–361.
7. BATEMAN, J. F., & PAPEZ, J. W. Significance of the thalamus in psychoses with reference to hallucinations and delusions. *J. clin. exp. Psychopath.*, 1951, **12**, 89–103.
8. BELLAK, L. A multiple factor psychosomatic theory of schizophrenia. *Psychiat. Quart.*, 1949, **23**, 738–755.
9. BELLAK, L. *Dementia praecox.* New York: Grune & Stratton, 1948.
10. BENDER, LAURETTA. Childhood schizophrenia. *Psychiat. Quart.*, 1953, **27**, 663–681.
11. BENDER, LAURETTA. Anxiety in disturbed children in P. Hoch, & J. Zubin, (Eds.) *Anxiety.* New York: Grune & Stratton, 1950.
12. BLEULER, E. *Dementia praecox or the group of schizophrenias.* New York: International Universities Press, 1950.
13. BLEULER, E. The physiogenic and psychogenic in schizophrenia. *Amer. J. Psychiat.*, 1930, **87**, 203–211.
14. BRACKBILL, G. A., & FINE, H. Schizophrenia and central nervous system pathology, in press.
15. BYCHOWSKI, G. Certain problems of schizophrenia in the light of cerebral pathology. *J. nerv. ment. Dis.*, 1935, **81**, 280–298.
16. BYCHOWSKI, G. Physiology of schizophrenic thinking. *J. nerv. ment. Dis.*, 1943, **98**, 368–386.
17. CLEGHORN, R. A. Endocrine influence on personality and behavior. In Milbank Memorial Fund, *The biology of mental health and disease.* New York: Hoeber, 1952.
18. DANZIGER, L., & KINDWALL, J. A. Thyroid therapy in some mental disorders. *Dis. nerv. Syst.*, 1953, **14**, 3–13.
19. DAVIS, PAULINE, & DAVIS, H. The electroencephalograms of psychotic patients. *Amer. J. Psychiat.*, 1939, **95**, 1007–1020.
20. DAVIS, PAULINE. Evaluation of the electroencephalograms of schizophrenic patients. *Amer. J. Psychiat.*, 1940, **96**, 851–860.
21. DEJONG, H. *Experimental catatonia.* Baltimore: Williams & Wilkins, 1945.
22. DESHON, J., RINKEL, M., & SOLOMON, D. Mental changes experimentally produced by LSD. *Psychiat. Quart.*, 1952, **26**, 33–53.
23. DOUST, J. W. L. Spectroscopic and photoelectric oximetry in schizophrenia and other psychiatric states. *J. ment. Sci.*, 1952, **98**, 143–160.
24. DUNLAP, C. B. The histopathology of dementia praecox. *Amer. J. Psychiat.*, 1924, **80**, 403–421.
25. EBAUGH, F., BARNACLE, C., & NEUBUERGER, K. Fatalities following electric convulsive therapy. *Arch. neurol. Psychiat.*, 1943, **49**, 107–119.
26. EICKHOFF, LOUISE F. W. The etiology of schizophrenia in childhood. *J. ment. Sci.*, 1952, **98**, 229–234.
27. ELVIDGE, A. R., & REED, G. E. Biopsy studies of cerebral pathologic changes in schizophrenia and manic depressive psychosis. *Arch. neurol. Psychiat.*, 1938, **40**, 227–268.
28. ERICKSON, M. The concomitance of organic and psychologic changes during marked improvement in schizophrenia. *Amer. J. Psychiat.*, 1934, **90**, 1349–1367.
29. FERRARO, A. Histopathological findings in two cases clinically diagnosed dementia praecox. *Amer. J. Psychiat.*, 1934, **90**, 883–903.
30. FERRARO, A. Pathological changes in the brain of a case clinically diagnosed schizophrenia. *J. neuropath. exper. Neurol.*, 1943, **2**, 84–94.
31. FINLEY, K., & CAMPBELL, M. Electroencephalography in schizophrenia. *Amer. J. Psychiat.*, 1941, **98**, 374–381.
32. FISCHER, R. Factors involved in drug-produced model psychoses. *J. ment. Sci.*, 1954, **100**, 623–632.
33. FREEMAN, W. Deficiency of catalitic iron in the brain in schizophrenia. *Arch. neurol. Psychiat.*, 1930, **24**, 300–310.
34. GIBBS, F. A., GIBBS, E. L., & LENNOX, W. G. The likeness of the cortical dysrhythmias of schizophrenia and psychomotor epilepsy. *Amer. J. Psychiat.*, 1938, **95**, 255–269.
35. GRINKER, R. & SEROTA, H. Electroencephalographic studies of corticohypothalamic relations in schizophrenia. *Amer. J. Psychiat.*, 1941, **98**, 385–392
36. HAMISTER, R. Effects of chlorpromazine on maze performance of rats. (In progress.)
37. HANFMANN, EUGENIA, & KASANIN, J. Conceptual thinking in schizophrenia. *Nerv. ment. dis. Monogr.*, 1942, No. 67.
38. HEATH, R. G. *Studies in schizophrenia.* Cambridge: Harvard Univer. Press., **1954**.

39. HOAGLAND, H. Stress and the adrenal cortex with special reference to potassium metabolism. *Psychosom. Med.*, 1950, **12**, 142–148.
40. HOAGLAND, H., & PINCUS, G. The nature of the adrenal stress response in schizophrenic men. *J. nerv. ment. Dis.*, 1950, **111**, 434–439.
41. HOAGLAND, H. Metabolic and physiologic disturbances in the psychoses. In Milbank Memorial Fund, *The biology of mental health and disease*. New York: Hoeber, 1952.
42. HOAGLAND, H., *et al.* Study of adrenocortical physiology in normal and schizophrenic men. *Arch. neurol. Psychiat.*, 1953, **69**, 470–485.
43. HOCH, P. H., CATTELL, J. P., & PENNES, H. H. Effects of drugs; Theoretical considerations from a psychological viewpoint. *Amer. J. Psychiat.*, 1952, **108**, 585–589.
44. HOFFER, A., OSMUND, H., & SMYTHIES, J. Schizophrenia: A new approach. II. Result of a year's research. *J. ment. Sci.*, 1954, **100**, 29–45.
45. HOLT, E. K., & TEDESCHI, C. Cerebral patchy demyelination. *J. neuropath. exper. Neurol.*, 1943, **2**, 306–314.
46. HOSKINS, R. G. *The biology of schizophrenia*. New York: Norton, 1946.
47. HYDE, R., VONMERING, O., & MORIMOTO, K. Hostility in the lysergic psychosis. *J. nerv. ment. Dis.*, 1953, **118**, 266–267. (Abstract)
48. JASPER, H., FITZPATRICK, C., & SOLOMON, P. Analogies and opposites in schizophrenia and epilepsy. *Amer. J. Psychiat.*, 1939, **95**, 835–851.
49. KANTOR, R. E., WALLNER, J., & WINDER, C. L. Process and reactive schizophrenia. *J. consult. Psychol.*, 1953, **17**, 157–162.
50. KENNARD, MARGARET, & LEVY, S. The meaning of the abnormal electroencephalogram in schizophrenia. *J. nerv. ment. Dis.*, 1952, **116**, 413–422.
51. KIRSCHBAUM, W. R., & HEILBRUNN, G. Biopsies of the brain of schizophrenic patients and experimental animals. *Arch. neurol. Psychiat.*, 1944, **51**, 155–162.
52. MAYER-GROSS, W., MCADAM, W., & WALKER, J. W. Further observations on the effects of lysergic acid diethylamide. *J. ment. Sci.*, 1953, **99**, 804–808.
53. MEADOW, A., & FUNKENSTEIN, D. The relationship of abstract thinking to the autonomic nervous system in schizophrenia. In P. H. Hoch & J. Zubin, (Eds.) *Relation of psychological tests to psychiatry*. New York: Grune & Stratton, 1952.
54. MINER, W. (Ed). Reserpine (serpasil) and other alkaloids of rauwolfia serpentina; chemistry pharmacology and clinical application. *Ann. N. Y. Acad. Sci.*, 1954, **59**, 1–140.
55. MOORE, M., NATHAN, D., ELLIOTT, A., & LAUBACH, C. Encephalographic studies in schizophrenia (dementia praecox). *Amer. J. Psychiat.*, 1932, **12**, 801–810.
56. NIELSEN, J. M. The basic pathology of schizophrenia. *J. nerv. ment. Dis.*, 1948, **107**, 340–357.
57. PAPEZ, J. W. Neuronal disease associated with intracytoplasmic inclusion bodies. *Arch. neurol. Psychiat.*, 1944, **52**, 217–229.
58. PAPEZ, J. W. Inclusion bodies associated with destruction of nerve cells in scrub typhus, psychoses and multiple sclerosis. *J. nerv. ment. Dis.*, 1948, **108**, 431–434.
59. PAPEZ, J. W., & BATEMAN, J. F. Cytological changes in nerve cells in dementia praecox. *J. nerv. ment. Dis.*, 1949, **110**, 425–437.
60. PAPEZ, J. W., & BATEMAN, J. F. Cytologic changes in cells of thalamic nuclei in senile, paranoid, and manic psychoses. *J. nerv. ment. Dis.*, 1950, **112**, 401–403.
61. PAPEZ, J. W., & BATEMAN, J. F. Changes in nervous tissue and study of living organisms in mental disease. *J. nerv. ment. Dis.*, 1951, **114**, 400–412.
62. PAPEZ, J. W. Form of living organisms in psychotic patients. *J. nerv. ment. Dis.*, 1952, **116**, 375–391.
63. PAPEZ, J. & PAPEZ, PEARL. The hypophysis cerebri in psychosis. *J. nerv. ment. Dis.*, 1954, **119**, 326–343.
64. PENNES, H. Clinical reactions of schizophrenics to sodium amytal, pervitin hydrochloride, mescaline sulphate, and D lysergic acid diethylamide (LSD 25). *J. nerv. ment. Dis.*, 1954, **119**, 95–112.
65. PIOTROWSKI, Z. A. The Rorschach inkblot method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.
66. POLATIN, P., EISENSTEIN, V., & BARRERA, E. Organic psychoses simulating dementia praecox. *Psychiat. Quart.*, 1944, **18**, 391–412.
67. RINKEL, M., DESHON, H., HYDE, R., & SOLOMON, H. Experimental schizophrenia-like symptoms. *Amer. J. Psychiat.*, 1952, **108**, 572–578.

68. ROIZAN, L., MORIARITY, J., & WEIL, A. Schizophrenic reaction syndrome in course of acute demyelination of central nervous system. *Arch. Neurol. Psychiat.*, 1945, **53**, 202–211.
69. ROWLAND, L. P., & METTLER, F. A. Cell concentration and laminar thickness in frontal cortex of psychotic patients: Studies on cortex removed at operation. *J. comp. Neurol.*, 1949, **90**, 255–280.
70. RUBIN, M. A. Electroencephalography in the psychoses. *Amer. J. Psychiat.*, 1940, **96**, 861–875.
71. RUPP, C., & WILSON, G. General pathologic findings associated with cases of so-called functional psychoses. *J. nerv. ment. Dis.*, 1949, **110**, 419–424.
72. SCHADERWALD, M. Cerebral cortical "inclusions" in psychoses. *J. comp. Neurol.*, 1953, **99**, 23–41.
73. SLOANE, B., & DOUST, J. Psychophysiological investigations in experimental psychoses: Results of the exhibition of d-lysergic acid diethylamide to psychiatric patients. *J. ment. Sci.*, 1954, **100**, 129–144.
74. SPIELMEYER, W. The problem of the anatomy of schizophrenia. *J. nerv. ment. Dis.*, 1930, **72**, 241–250.
75. STEIN, R. J., & ZIEGLER, L. A comparison of the thalamus in dementia praecox and manic depressive brains: A biometric analysis. *J. nerv. ment. Dis.*, 1939, **90**, 709–729.
76. TRAVIS, L. E., & MALAMUD, W. Brain potentials from normal subjects, stutterers and schizophrenic patients. *Amer. J. Psychiat.*, 1937, **93**, 929–932.
77. TURNER, W., & LOWINGER, L. The correlation of pre-electroshock electroencephalogram and therapeutic result in schizophrenia. *Amer. J. Psychiat.*, 1945, **102**, 299–300.
78. VAN DER HORST, L. Histopathology of clinically diagnosed schizophrenic psychosis or schizophrenia-like psychosis of unknown origin. *Acta. psychiat. Kbh.*, 1953, **28**, 191–199.
79. WEINSTEIN, M. Histopathological changes in the brain in schizophrenia. *Arch. neurol. Psychiat.*, 1954, **71**, 539–553.
80. WINKELMAN, N. W., & BOOK, M. H. Observations on the histopathology of schizophrenia. I. The cortex. *Amer. J. Psychiat.*, 1949, **105**, 889–896.
81. WOLF, A., & COWEN, D. Histopathology of schizophrenia and other psychoses of unknown origin. In Millbank Memorial Fund. *The biology of mental health and disease.* New York: Hoeber, 1952.

Received April 21, 1955.

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 3, 1956

EXPERIMENTAL EVALUATIONS OF ROLE PLAYING[1]

JOHN H. MANN
Columbia University

In recent years a number of articles have been written on the use of role playing in education, personality assessment, role training, and psychotherapy. Yet, in spite of the widespread popularity of role playing, comparatively few studies have been reported in which role playing was evaluated experimentally.

The present article reviews such experimental studies in an attempt to establish the degree to which present practice is justified by experimental evidence, and to indicate the problems toward which future experimentation might most profitably be directed.

A role-playing situation is here defined as a situation in which an individual is *explicitly* asked to take a role not normally his own, or if his own in a setting not normal for the enactment of the role.

Examples of taking a role not one's own are typically found in assessment procedures. For example:

> You are a twelve-year-old boy without brothers and sisters. All year long you have been looking forward to having your first experience in boy scout camp which runs for two months this summer. Camp begins next week. You are starting to pack as your father enters the room (7, p. 374).

Taking one's own role in an unusual setting typically occurs in the context of a procedure designed to produce personality change. In such situations the individual is asked to portray scenes from his own life before a group of other people. The individual is acting his own role but the setting is unusual. For example:

> We would like to see how you normally handle a customer when he comes in your store. Suppose Mr. X plays one of your customers. He has just asked you the price of a certain article. What do you do?

Studies of role playing can conveniently be grouped in accordance with the two definitions considered above. That is, they can be considered to treat role playing either as a method for assessing personality or as a method for producing personality and behavioral changes. This classification will be adopted in the review which follows.[2]

ROLE PLAYING AS AN ASSESSMENT PROCEDURE

Early work in this area was strictly qualitative. In 1945, for example, Moreno and Moreno (17) used role playing to study the social stereotypes of children. Other examples, summarized by Eaton (9), occurred during World War II. At this time use was made of role playing in the assessment and selection of personnel. The studies of the OSS (18) and the proposed standardized assessment procedures of Bronfenbrenner and Newcomb (7) are outstanding examples of the qualitative work done during this period. Interest in quantitative measurement did not

[1] I am indebted to Dr. Edgar Borgatta and Dr. Leonard Cottrell of the Russell Sage Foundation for reading this manuscript and providing many helpful suggestions.

[2] There have been a number of studies in which the experimenters have consciously played a role as part of the experimental procedure. These studies are not included in this review for two reasons. First, the experimental subjects themselves do not role-play. Second, role playing is used strictly as an experimental technique for fabricating reality, and is not itself the object of study.

occur in this area until experimenters became concerned with the reliability and validity of their measuring instruments.

Reliability

Probably the first to test the reliability of ratings of role-playing performance were Rotter and Wickens (**20**). They asked observers to rate subjects placed in two standard role-playing situations on the single characteristic of "social aggressiveness." During one situation the subject was observed by a panel of four observers. During the second situation he was observed by a different panel of observers. There were also three observers who watched his behavior in both situations. Under these conditions a mean interrater reliability coefficient of .71 was obtained for a panel observing a single situation (N of 48). The correlation between the mean ratings of the members of the two panels on the behavior of the 48 subjects over two situations was .55, while the mean ratings of the three observers who watched the subjects in both situations had a reliability coefficient of .77. The .55 correlation may be interpreted not only as rater reliability but as subject consistency from situation to situation. The larger .77 correlation may result in part from a halo effect.

Stanton and Litwak (**22**), in a study devoted to evolving a test of autonomy, asked five judges to record the number of times they heard examples of "nonideal" behavior on 11 tape recordings of role-playing situations.[3] In this way they obtained a mean reliability coefficient of .90 (rho).

Moldawsky (**16**), also using tape recordings of role-playing situations, was able to obtain an interrater reliability coefficient of .89 for ratings of "rigidity" of behavior.

Kelly and Fiske (**13**), interested in predicting success in clinical psychology, asked judges to rate subjects in role-playing situations on characteristics which were unrelated to the immediate role-playing enactment. Among these characteristics were research ability, therapeutic ability, diagnostic ability, and overall suitability for the profession of clinical psychology. Interjudge reliability for these characteristics was found to be .45, .44, .40, and .51 respectively. In this study, however, not only was the reliability of the observer's judgment being measured but also the degree of similarity between his conception of the abilities required for successful performance in research, therapy, and diagnosis and that of the other judges. In the single case in which the individual judges' ratings were averaged and corrected by the Spearman-Brown formula, the reliability for the given variable rose from .51 to .76.

Borgatta (**2**) has studied the reliability of "real" situations and role-playing situations. In his study the observation system (Bales Interaction Analysis) was held constant and the reliability of the categories in the system was measured within the types of situations. It was found that the kinds of behavior which an individual initiates and receives are about as consistent in role playing as in "actual" behavior.

Almost any system for observing human interaction which has been devised can be used for observing role playing. The reliability associated with such systems is therefore related to the present discussion, but here we report only the findings derived from interaction situations

[3] Stanton and Litwak equate nonideal behavior with lack of autonomy. This may be a dubious equation, but it does not affect the present discussion.

specifically labeled role playing. In summary, it can be said that the above studies tend to confirm the possibility of establishing reliable systems of observing role-playing behavior.

Validity

Studies of validity can be ordered in relation to the nearness of the testing situation to the validating situation. The following discussion is ordered according to this principle, starting with studies in which the two are adjacent and extending to studies in which the testing and validating situations are only tenuously related.

Borgatta (**2**, **3**, **4**, **5**, **6**) in a series of reports has described the results of a study in which subjects alternately role-played and behaved "actually" in an experimental setting. The subjects of this experiment were assigned to 166 three-man groups and rotated in such a way as to ensure that each subject would have equal experience with an array of persons.

Perhaps the most striking finding of this study in relation to the present context was that, though the interaction profiles varied greatly between the role-playing and actual situation, the performance of group members remained remarkably constant in relation to each other. Specifically the correlation between total rate of interaction initiated and received and rate of positive emotional responses initiated and received over the two situations was .76, .66, .58, and .58 respectively. (A correlation of .17 was significant at the .05 level.) In addition, a close parallelism was found between role playing and actual behavior over all the Bales Interaction Categories except that of "asking for suggestions" which was a null variable for the actual behavior.

Certain differences between role playing and actual behavior were noted however. Role playing was characterized by lack of tension and inhibition, and by higher rates of asking for and giving opinion in comparison to the actual behavior in which more emphasis was placed on neutrality of feeling and task orientation. It is also interesting to note that leadership self-ratings made by subjects, and leadership ratings made by their associates appeared to correlate more highly with their performances in role playing than with their performances in the actual behavior situation.

In a further analysis of these data Borgatta (**6**) compared role playing and actual behavior with behavior on a paper-and-pencil projective test (Conversation Study). The test consisted of a leaflet in which ten plates with various three-man situations were drawn. The subject was asked to write down what each of the men was saying. This imagined conversation was then analyzed by the Bales Interaction Categories in the same manner as the role playing and the actual behavior. From these data a correlation of .3 was obtained between the total rate of response on the projective test and that of either the role playing or the actual behavior. Role playing and actual behavior, as previously mentioned, were highly correlated.

This finding was similar to one obtained in an earlier experiment by Borgatta (**1**). In that pilot experiment the comparison was made between a subject's response to the Rosenzweig P-F Study when given normally (in written form), when used as the basis for a series of role-playing incidents, and as the basis for an analogous series of apparently unplanned real-life incidents. In this earlier experiment, in which 78 subjects were utilized, the findings were

not clear because of the low reliability associated with the scoring system which distributed 24 scores over 11 categories.

To clarify further the relationship among these three types of situations Borgatta (6) subjected his data from the later study to a factor analysis. Of five relevant factors found, two, *task ability* and *emotional assertiveness*, showed close parallelism over role playing and actual behavior and some parallelism over the projective test. A third factor, *military adjustment*, showed some parallelism over role playing and actual behavior, with a suggestion of parallelism over the projective test. In a fourth factor, *task supportiveness*, parallelism was found over role playing and actual situation only. Finally in a fifth factor, *emotional group supportiveness*, no parallelism was found.

In his work Borgatta measured actual and role-playing behavior which occurred in the same experimental setting. Stanton and Litwak (22) expanded the distance between these types of behaviors by studying the relationship between role playing occurring in an experimental setting and actual behavior occurring in a natural setting. They were interested in studying autonomy as an aspect of stress-tolerance. They used a test which consisted of three standard stress situations. Raters were asked to note the number of nonideal behaviors which occurred during the enactment. The validity of the test was determined by correlating the measure of autonomy so obtained using a group of foster parents as subjects with ratings of autonomy made by social caseworkers who had considerable contact with the foster parents. A correlation of .82 was obtained. A correlation of only .48 was found between the caseworker's rating and the rating of the homefinder, who interviewed the foster parents in their home. In a small class of eight students, a correlation of .79 was found between ratings of autonomy determined from role playing and ratings made by close friends.

It is also interesting to note that while raters who saw only role-playing scenes referring to autonomy were able to make ratings that correlated .82 with caseworkers' ratings, raters who saw other role-playing scenes in addition, were able to make ratings which correled only .40 with caseworkers' ratings. This suggests that the additional information which raters received contaminated, rather than clarified, the rater's perception of the role-playing situation.

Kelly and Fiske (13) attempted to use role-playing behavior as a basis for making predictions of future success in clinical psychology. In their large-scale assessment program they utilized a battery consisting of standardized paper-and-pencil tests, projective tests, interviews, written background material, and situational (role-playing) tests. Two kinds of validation are provided by these data. First, one can consider the correlation between the situational test rating and the over-all rating obtained by the use of the whole test battery which, in this case, is the validating criterion. It was found that observers seeing only the situational tests could make ratings as good as those made by other observers who read the subject's background credentials and conducted a preliminary interview with him, or who had seen his background credentials, his objective and projective test results, and had read his autobiography. Second, one can consider the correlation between the situational test rating and the criteria utilized to establish later successful performance in clinical psychology.

The correlation between the over-all score based on the test battery and the validation criterion was .34. The correlation between the situational tests and the validating criterion was .20. This is not impressive since an equivalent correlation was obtained by using either the credential file or a standard intelligence test score. This low correlation may reflect the fact that the role-playing situations used were unrelated to the validating criterion in content.

In summary, it can be said that there is some evidence which indicates that valid predictions of interpersonal behavior can be made from role-playing assessment procedures. The actual development of such procedures has, to date, been very slow and consequently the problems involved have hardly been explored.

Role Playing as a Method for Producing Personality Change

Role playing has been used to produce personality and behavioral change in a wide variety of settings varying from leadership and teacher training to psychotherapy with neurotics and psychotics. Comparatively little experimental work has been reported, though application has been extensive.

Harrow (**10**) studied the effect of role playing on 20 schizophrenics organized into two groups that met 25 times over a two-month period. A third group of schizophrenics was used as a control. Each individual was given the Rorschach, the MAPS, and a special role test before and after the role-playing series. Although statistically significant results were not obtained, a group of judges looking at the before-after data found evidence of the development of more realistic perception of and interest in the outside world as well as an increased ability to handle personal problems.

Jones and Peters (**12**) in a similar experiment studied the effect of role playing on a group of twelve schizophrenics and a group of eight lobotomy patients. Another group of schizophrenics was used as a control. The role-playing sessions were held once a week for four months. Before-after evaluation was made using the Porteus Maze, the Mirror-Tracing Test, Gardner Behavior Chart, Rorschach, and the Draw-A-Man Test. Again, although statistically significant results were not obtained, it was found that all tests except the Draw-A-Man Test indicated improvement.

Sause (**21**) has investigated the effect of role playing on a group of 15 normal student teachers. At the beginning of each role-playing session the group members selected a problem related to student-teacher interaction as the focus of the session. They then each wrote down their solution to the problem. Following this they role-played the problem and again wrote down their solution. Finally they rated their satisfaction with the meeting. Four judges rating the written solutions before and after a given role-playing session were practically unanimous in finding an improvement in the solutions after role playing. The judges agreed with each other 80 per cent of the time. It was also found that the quality of the solutions reached before role playing improved as the sessions progressed. Eighty-seven per cent of the student teachers were rated as having better solutions before role-playing at the twelfth meeting than at the third, indication that a transfer of training may have occurred.

Maier (**15**), also working with normal groups, has tested the effect of a specialized form of role playing (Multiple Role Playing) in helping groups

reach satisfactory solutions to common problems. Using this technique he found that 42 out of 44 groups reached a solution, and that only four per cent of the group members were dissatisfied with the solution reached.

Because of either small sample size, design, or lack of underlying relationship, none of the four preceding studies provide statistically significant results. In addition, two of them fail to utilize a control group. It is best, therefore, to consider them as exploratory in nature.

A more satisfactory study from the methodological viewpoint was that of Janis and King (**11**). In their design they utilized 90 college students divided into groups of three. Each student gave a short talk to the other two on an assigned topic. Each student, therefore, spoke once and listened twice. The topics were assigned in such a way that the student speaking presented a point of view contrary to his own as determined by a previous opinion questionnaire. At the end of each speaking session (i.e., role-playing session) the opinions of the listeners and the speaker were reassessed.

The results indicated that, for two of the speaking topics used, there was significantly more change of attitude in the speaker than in the listener. For the third topic the change was not significant, but the speakers expressed greater confidence in the sureness of their opinions than the listeners. It was also found that individuals who improvised most and who were satisfied with their performance as speakers underwent the greatest amount of opinion change. There are two difficulties in generalizing the results of this study. First, speaking on an assigned topic before two other people is a very limited form of role playing. It involved only a one-way interaction between the active speaker and the passive audience. The typical role-playing situation consists of at least two-way active interaction. Second, the opinions changed were unimportant. They dealt with the number of movie theaters that would exist in three years, the size of the probable meat supply in the following year, and the length of time that would be required to find a cure for the common cold. It is probable that none of these topics was of central importance to the students tested. If this had not been the case the students might have shown greater resistance to change.

The results obtained by Janis and King were replicated by Rosenberg (**19**) under less artificial conditions. Using three standard role-playing situations with three groups of 15 persons each, group members were rotated among three experimental positions, *role players*, *identifiers* (instructed to identify with the role players), and *active observers* (instructed to watch the action as objectively as possible). A special group of control observers was assigned simply to watch the action. The role-playing situation involved such topics as labor-management relations, guidance problems, etc. After each role-playing session a questionnaire was given to all group members. This was followed by a carefully observed nondirective group discussion about the role playing. Using data so gathered it was found that the role players were most susceptible to change. They were also found to be the most productive in the discussion after the role playing, to have certain emotional biases of judgment and perception, and a tendency to support their own actions in the role playing. The identifiers were found to be most critical, almost as productive as the role players during the dis-

cussion, and to have an awareness of the dynamics of the interaction process second only to the role player. The active observer was found to be uninvolved and therefore unemotional with reference to the enactment. He was as critical as the identifier but relatively unaware of the dynamics of the interaction process. Finally, the control observer was found to be indifferent, unproductive during the discussion, and relatively unaware of the interaction dynamics. An interesting implication of this study is that understanding of the dynamics of interaction processes is associated with involvement in those processes. It is also interesting to note that individual involvement was found to be a function of position assignment.

While Rosenberg was concerned with the differences induced by experimental position within the role-playing situation, she did not consider differences between individuals in the same position. These differences have been studied by Luszki (**14**) who compared observers who were good at understanding the interaction dynamics in a role-playing scene with those who were not. She found that good observers, as distinguished from poor ones, (*a*) were good at determining how others looked at themselves, (*b*) were good at determining how others looked at them, (*c*) were well adjusted, (*d*) evaluated themselves as others evaluated them, and (*e*) were consistently and favorably perceived by others. Further it was found that persons who were good judges of how others felt about them were able to identify most successfully with role players. Persons who were good judges of how others felt about themselves were found to be good at observing the actions of role players critically.

Brown (**8**) has explored the effect of the role-playing situation on the success of the role player's performance. He did this by asking his subjects to reverse roles with each other and having them evaluate how well the other person succeeded in portraying them. It was found that success was inversely related to the realism of the role-playing situation. The more unusual the situation in which the role was portrayed, the better was the portrayal.

In summary, the findings in this area are sketchy and essentially suggestive in nature. The evidence indicates that role playing may produce behavioral and personality change, but this has certainly not been established with any degree of confidence. The evidence also suggests that the effect of role playing on the individual is related to his personality and to his position in the role-playing situation. In addition, the success of the role player may be related to the setting of action in the role-playing situation.

Summary

Perhaps the most striking impression to be gained from a review of the experimental studies of role playing is their scarcity. Except for the few studies reported here, effort among practitioners of role playing seems to have been devoted to the exploitation of its various applications.

With reference to personality assessment, there is some sound evidence for believing that reliable and valid role-playing tests can be developed. Such tests are, however, still in their infancy, and it is toward the development of these tests that future effort might profitably be directed.

With reference to personality change, it can only be said that there is as yet little supportive evidence. The studies which do exist suggest the possibility of such change and in-

reach satisfactory solutions to common problems. Using this technique he found that 42 out of 44 groups reached a solution, and that only four per cent of the group members were dissatisfied with the solution reached.

Because of either small sample size, design, or lack of underlying relationship, none of the four preceding studies provide statistically significant results. In addition, two of them fail to utilize a control group. It is best, therefore, to consider them as exploratory in nature.

A more satisfactory study from the methodological viewpoint was that of Janis and King (**11**). In their design they utilized 90 college students divided into groups of three. Each student gave a short talk to the other two on an assigned topic. Each student, therefore, spoke once and listened twice. The topics were assigned in such a way that the student speaking presented a point of view contrary to his own as determined by a previous opinion questionnaire. At the end of each speaking session (i.e., role-playing session) the opinions of the listeners and the speaker were reassessed.

The results indicated that, for two of the speaking topics used, there was significantly more change of attitude in the speaker than in the listener. For the third topic the change was not significant, but the speakers expressed greater confidence in the sureness of their opinions than the listeners. It was also found that individuals who improvised most and who were satisfied with their performance as speakers underwent the greatest amount of opinion change. There are two difficulties in generalizing the results of this study. First, speaking on an assigned topic before two other people is a very limited form of role playing. It involved only a one-way interaction between the active speaker and the passive audience. The typical role-playing situation consists of at least two-way active interaction. Second, the opinions changed were unimportant. They dealt with the number of movie theaters that would exist in three years, the size of the probable meat supply in the following year, and the length of time that would be required to find a cure for the common cold. It is probable that none of these topics was of central importance to the students tested. If this had not been the case the students might have shown greater resistance to change.

The results obtained by Janis and King were replicated by Rosenberg (**19**) under less artificial conditions. Using three standard role-playing situations with three groups of 15 persons each, group members were rotated among three experimental positions, *role players*, *identifiers* (instructed to identify with the role players), and *active observers* (instructed to watch the action as objectively as possible). A special group of control observers was assigned simply to watch the action. The role-playing situation involved such topics as labor-management relations, guidance problems, etc. After each role-playing session a questionnaire was given to all group members. This was followed by a carefully observed nondirective group discussion about the role playing. Using data so gathered it was found that the role players were most susceptible to change. They were also found to be the most productive in the discussion after the role playing, to have certain emotional biases of judgment and perception, and a tendency to support their own actions in the role playing. The identifiers were found to be most critical, almost as productive as the role players during the dis-

cussion, and to have an awareness of the dynamics of the interaction process second only to the role player. The active observer was found to be uninvolved and therefore unemotional with reference to the enactment. He was as critical as the identifier but relatively unaware of the dynamics of the interaction process. Finally, the control observer was found to be indifferent, unproductive during the discussion, and relatively unaware of the interaction dynamics. An interesting implication of this study is that understanding of the dynamics of interaction processes is associated with involvement in those processes. It is also interesting to note that individual involvement was found to be a function of position assignment.

While Rosenberg was concerned with the differences induced by experimental position within the role-playing situation, she did not consider differences between individuals in the same position. These differences have been studied by Luszki (**14**) who compared observers who were good at understanding the interaction dynamics in a role-playing scene with those who were not. She found that good observers, as distinguished from poor ones, (*a*) were good at determining how others looked at themselves, (*b*) were good at determining how others looked at them, (*c*) were well adjusted, (*d*) evaluated themselves as others evaluated them, and (*e*) were consistently and favorably perceived by others. Further it was found that persons who were good judges of how others felt about them were able to identify most successfully with role players. Persons who were good judges of how others felt about themselves were found to be good at observing the actions of role players critically.

Brown (**8**) has explored the effect of the role-playing situation on the success of the role player's performance. He did this by asking his subjects to reverse roles with each other and having them evaluate how well the other person succeeded in portraying them. It was found that success was inversely related to the realism of the role-playing situation. The more unusual the situation in which the role was portrayed, the better was the portrayal.

In summary, the findings in this area are sketchy and essentially suggestive in nature. The evidence indicates that role playing may produce behavioral and personality change, but this has certainly not been established with any degree of confidence. The evidence also suggests that the effect of role playing on the individual is related to his personality and to his position in the role-playing situation. In addition, the success of the role player may be related to the setting of action in the role-playing situation.

Summary

Perhaps the most striking impression to be gained from a review of the experimental studies of role playing is their scarcity. Except for the few studies reported here, effort among practitioners of role playing seems to have been devoted to the exploitation of its various applications.

With reference to personality assessment, there is some sound evidence for believing that reliable and valid role-playing tests can be developed. Such tests are, however, still in their infancy, and it is toward the development of these tests that future effort might profitably be directed.

With reference to personality change, it can only be said that there is as yet little supportive evidence. The studies which do exist suggest the possibility of such change and in-

dicate a few of the relevant variables that may be involved. Much effort is needed in the exploration of role playing as a method for producing change. If the possibility of such change were established a number of related topics become pertinent. For example: What personality characteristics are most affected by role playing? Are all role-playing techniques equally effective? Is role playing less effective in producing change if it is preplanned? What is the effect of the audience on the role players? These questions, and others which will readily occur to the reader, are open for experimentation. But they all rest on the assumption that role playing can produce personality change, and this has not as yet been conclusively demonstrated.

BIBLIOGRAPHY

1. BORGATTA, E. F. An analysis of three levels of response: An approach to some relationships among dimensions of personality. *Sociometry*, 1951, **14**, 267–316.
2. BORGATTA, E. F., & BALES, R. F. The consistency of subject behavior and the reliability of scoring in interaction process analysis. *Amer. sociol. Rev.*, 1953, **18**, 566–569.
3. BORGATTA, E. F., & BALES, R. F. Task and accumulation of experience as factors in the interaction of small groups. *Sociometry*, 1953, **16**, 239–252.
4. BORGATTA, E. F., & BALES, R. F. Interaction of individuals in reconstituted groups. *Sociometry*, 1953, **16**, 302–320.
5. BORGATTA, E. F. Analysis of social interaction and sociometric perception. *Sociometry*, 1954, **17**, 7–32.
6. BORGATTA, E. F. Analysis of social interaction: actual, role playing and projective. *J. abnorm. soc. Psychol.*, 1955, **51**, 394–405.
7. BRONFENBRENNER, U., & NEWCOMB, T.M. Improvisations—an application of psychodrama in personality diagnosis. *Sociatry*, 1948, **1**, 367–382.
8. BROWN, J. C. An experiment in role-taking. *Amer. sociol. Review*, 1952, **17**, 587–597.
9. EATON, J. W. Experiments in testing for leadership. *Amer. J. Sociol.*, 1947, **52**, 523–535.
10. HARROW, GERTRUDE. Psychodrama group therapy. *Group Psychother.*, 1952, **5**, 120–170.
11. JANIS, J. L., & KING, B. T. The influence of role playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.
12. JONES, F. D., & PETERS, A. N. An experimental evaluation of group psychotherapy. *J. abnorm. soc. Psychol.*, 1952, **47**, 345–353.
13. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
14. LUSZKI, M. B. Empathic ability and social perception. Unpublished doctor's dissertation, Univer. of Michigan, 1951.
15. MAIER, N. R. F. *Principles of human relations.* New York: Wiley, 1952.
16. MOLDAWSKY, S. An empirical validation of a rigidity scale against a criterion of rigidity in an interpersonal situation. *Sociometry*, 1951, **14**, 153–174.
17. MORENO, F. B., & MORENO, J. L. Role tests and role diagrams of children. In *Group psychotherapy.* New York: Beacon House, 1945.
18. OSS ASSESSMENT STAFF. *The assessment of men.* New York: Rinehart, 1948.
19. ROSENBERG, PEARL P. An experimental analysis of psychodrama. Unpublished doctor's dissertation, Radcliffe Coll., 1952.
20. ROTTER, J. B., & WICKENS, D. D. The consistency and generality of ratings of "social aggressiveness" made from observations of role playing situations. *J. consult. Psychol.*, 1948, **12**, 234–239.
21. SAUSE, L. M. Role playing the selected problems of student teachers. Unpublished doctor's dissertation, Teachers College, Columbia Univer., 1954.
22. STANTON, H., & LITWAK, E. Towards the development of a short form test of interpersonal competence. Unpublished manuscript.

Received August 4, 1955

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 3, 1956

RESEARCH WITH THE WECHSLER-BELLEVUE INTELLIGENCE SCALE: 1950–1955[1]

WILSON H. GUERTIN
Higher Teachers College, Baghdad, Iraq
GEORGE H. FRANK
Topeka State Hospital
AND ALBERT I. RABIN
Michigan State University

Despite the somewhat pessimistically worded summary in our last review of the W-B[2] (**97**), the avalanche of research studies with this instrument has abated very little. The bibliography that accumulated attests to that. Moreover, this fact is a testament to the vitality of the instrument and to the tenacity of its numerous users, especially as a medium of research. The studies, again, cover the entire range of interests, with a weakening in some areas and greater emphasis in others. Thus, the issue of "short forms" has been nearly exhausted whereas sex differences is almost an entirely new topic of investigation. Generally, more attempts have been made to examine the measure of intelligence. These investigations and previous research have been instrumental in producing a larger number of studies characterized by greater care and sophistication.

[1] Through July 1955.

[2] The abbreviation, W-B, will be used throughout to indicate the Wechsler Bellevue Intelligence Scale (Form I). The names of the subtests also appear in abbreviated form throughout the paper. The single letters *I*, *C*, *A*, *D*, *S*, and *V* stand for the verbal subtests: Information, Comprehension, Arithmetic, Digits, Similarities, and Vocabulary, respectively. The two-letter combinations *PA*, *PC*, *OA*, *BD*, and *DS* correspond to the following subtests: Picture Arrangement, Picture Completion, Object Assembly, Block Design, and Digit Symbol. *FS*, *V* IQ, and *P* IQ stand for Full Scale, Verbal IQ, and Performance IQ, respectively.

A review by Watson (**127**) appeared in the interval since our last review. Our attempt will, again, be to make an exhaustive and critical examination of the studies in this area. Many of these would probably not have been published had authors and editors heeded the recommendations of our previous review. We have taken care to prepare a complete bibliography so that taken with the earlier two reviews (**96**, **97**) the literature should be well covered.

The organization of the present review resembles closely that of the previous one. The first half deals with the W-B as a test of general intelligence, whereas the second half of the article is devoted to studies with sundry classifications of deviant and disordered subjects.

THE WECHSLER-BELLEVUE AS A TEST OF GENERAL INTELLIGENCE

Studies of Reliability

Recent interest in practice effects and the dependability of subtest scores has filled the previously observed lacunae of reliability data on the W-B. Most of the currently reviewed studies present reliability coefficients for the subtests as well as for the three IQ scales.

Test-retest agreement. The two studies by Steisel (**117**, **118**) investigated the effect of practice on the W-B scores. Subjects were 17 matched pairs of college girls of su-

perior intelligence. The first of each pair was retested after 14 days while the retest interval for the second of each pair was 77 days. Practice effects on all three scales appeared for both groups. Mean increase in *FS* IQ was eight to nine points. The *P* IQ increases were approximately double those for the *V* IQ. Steisel (**117**) points out that " . . . in a retest situation the examiner should place more trust in the verbal scores . . . ." Differences in the retest scores after the short and long intervals revealed that only the score on the *A* subtest was a function of the recency of administration. The author concludes, "The findings in general corroborate those of previous studies in indicating that the significant gains in retest scores are maintained up to approximately three months . . . ." A similar study, but with mental defectives, is reported by Hays and Schneider (**56**). They employed different retest periods of two, four, six, and eight weeks. The results also suggest that the different retest intervals do not affect the IQ increases differentially. The W-B Form I *FS* IQ showed an increase of 7.6 points due to practice as compared with only 4.4 for the Form II.

Split-half reliability. When Webb and De Haan (**129**) reported higher split-half reliability coefficients for psychotics than for normals, they provoked much thought. Helmick (**57**) criticizes their failure to equalize the range of talent for the two groups. Webb (**128**) confidently replies that Helmick's suggestion of equalizing the ranges of talent would create a hybrid sample which was unreal and could never be representative of an actual population of people.

Perhaps less interesting but more important than these asides are the main features of the study. Webb and De Haan (**129**) employed 100 subjects to provide reliability coefficients for the subtests. All the subtest coefficients were statistically significantly different from zero for the paranoid subjects, while only *PA* failed to show significance with the normal half of the sample. The *V* subtest was the most reliable for both samples. Only two subtests showed much difference in reliability for the different subject sampling. These two subtests, *PA* and *PC*, were more reliable for the psychotics. Botwinick (**14**) utilized Webb and De Haan's normal group, a group of 50 older normals, and a group of 31 patients with mental disorders of the senium. Unfortunately sex could not be held constant for all three groups. Corrected split-half coefficients were calculated, and only *C* failed to differ significantly from zero. The author's findings were characteristic for this type of study. The coefficients were larger for the older subjects and were still greater for the senile disordered group. As the subtest variability increases due to aging or mental changes associated with illness, the reliability coefficients seem to increase in magnitude. The reliability of some subtests is not as high as might be desired. Thus, it is not surprising that the authors (**14**, **129**) are skeptical of employing diagnostic pattern analyses that are based upon the more unreliable of the subtests.

Summary. The additional research covered in the current section provides sufficient information to guide the clinician. The studies have been well designed and are fairly conclusive. The IQ scales seem reliable, although subject to practice effects. The practice-effect increment does not depend greatly upon the size of the test-retest interval provided the retesting occurs within three months. Practice effects are about twice as great on the performance

section. Split-half reliability coefficients for the various subtests show quite a range of values. Some of the smaller coefficients cast doubt upon the usefulness of diagnostic pattern analysis which is based upon the more unreliable subtests. Additional work in the area of reliability should be well designed and comprehensive.

Correlation With Other Tests

Other Wechsler scales. The W-B is compared with the Form II in a comprehensive study by Gerboth (**45**). Retest results disclosed that when Form II was administered first there was a significant increase in the subsequent W-B scales scores. The scores were uniformly higher by about five points for all scales. Surprisingly, Hays and Schneider (**56**) found this same elevation of W-B scores when they preadministered the Form II. This is surprising, because instead of normal superiors they employed mental defectives. For some reason preadministering the Form I does not elevate the subsequently obtained Form II scores nearly as much. A satisfactory explanation of this special effect has not been forthcoming.

Returning to Gerboth's study (**45**), the intercorrelations between the subtests suggest that the *I*, *C*, *A*, and *PA* subtests were least comparable between the W-B and the Form II. The correlation between *FS* IQ's was about .75. This is fairly high since the range of talent was only half that for an unrestricted sample.

Another Wechsler scale, The Wechsler Intelligence Scale for Children (WISC), was compared with the W-B by Delattre and Cole (**31**). With a good range of intelligence they found a correlation of .87 between the *FS* IQ's. The mean *FS* IQ was six points higher for the WISC. Intercorrelation of corresponding subtests of the two tests provides indices of comparability. *PA* was particularly low with an *r* of only .19 between the two tests. The rather moderate correlations provoke the authors to wisely invoke caution in applying W-B patterns and signs to WISC data. An almost identical design with a narrower range of intelligence is found in a more recent study by Knopf, Murfett, and Milstein (**68**). Their results provide good confirmation of the earlier findings of Delattre and Cole (**31**). They also emphasize the lack of comparability between corresponding subtests of the two tests. Vanderhost, Sloan, and Bensberg (**124**) report on the gross comparability of the W-B and the WISC for a mental defective group. Their experimental design neatly counterbalances extraneous factors. With a restricted range of IQ they report a correlation of .72 between *FS* IQ's. They found the mean WISC to be 1.6 IQ points higher than the W-B, a difference not statistically significant.

Other intelligence tests. Alderdice and Butler (**3**) found a correlation of .69 between the W-B *FS* and the Revised Stanford Binet (S-B), Form M. However, the correlation is somewhat attenuated by the narrow range of IQ's in the sample. Correlations between the W-B and the S-B are considerably higher when an appropriate range of talent is employed (**96**, **97**). As encountered in the past, the W-B IQ was found to be somewhat higher than that for the S-B (**3**). Duncan (**37**) presents a table for converting W-B vocabulary scores to S-B vocabulary scores, thereby eliminating the necessity of administering both vocabularies, when both tests are included in a battery.

The Army General Classification Test (AGCT) was compared with the W-B by Tamminen (**121**). A correlation of .91 was obtained between the

two tests when the sample was corrected for range of talent. The AGCT appears to be more closely related to the *V* IQ than to the *P* IQ of the W-B. The American Council on Education Test (ACE) correlated .61 with the W-B *FS* with a restricted range of IQ's in the sampling by Merrill and Heathers (**86**). Gerboth (**45**) reports a comparable correlation between the ACE and the W-B but also employed a narrow range of talent. A thesis by Smith (**109**) reports the relationship between the SRA Primary Mental Abilities Test but was not available for review.

Knott *et al.* (**69**) compared the W-B with the seven-test Kent Battery. The correlation with the *FS* IQ was .87, but this is difficult to interpret since the range of talent of the sample was not reported. Delp (**32**) compared only the EGY of the Kent Battery with the W-B. The correlation was .65, and the range of talent was only slightly reduced. Raven's Progressive Matrices correlated .55 with a short form of the W-B according to Levine and Iscoe (**71**). They state, "The Progressive Matrices appeared to tap areas of intelligence most closely related to *BD* and not significantly related to *C*." In another study Levine and Iscoe (**72**) found the correlation between the W-B and the Matrices " . . . not high enough for individual prediction . . . ," but this was with a deaf sample. Desai (**33**) compared the Matrices against just the *V* IQ of the W-B. He obtained an *r* of .65 which was corrected for attenuation. Allen, Thornton, and Stenger (**5**) found the surprisingly high correlation of .86 between the Ammons Picture Vocabulary and the W-B. Rubin (**101**) presents the results of his comparison of the W-B and quantitative features of the H-T-P. With an appropriate range of talent he obtained a correlation of .67. The mean IQ's were rather comparable.

Other measures. Storrs (**119**) reports interesting relationships between the W-B scales and the General Aptitude Test Battery (GATB). The Battery showed the interest, verbal, and numerical factors to be related to the *V* IQ. The test-factors related to *P* IQ were spatial, form perception, clinical perception, aiming, and motor speed. A correlation of .58 was found for the relationship between the total score for the Battery and the W-B *FS*. The correlation with the *P* section was slightly higher than with the *FS*.

Interest in the correspondence between W-B and intellectual aspects of Rorschach performance is encountered in the work of Abrams (**1**). A multiple-regression equation is presented which predicts IQ score. The Rorschach elements involved are $F+\%$, *M*, *W*, and *R*. The multiple *R* is .53.

Only two studies compared W-B scores with criteria other than tests. Gerboth (**45**) found that both forms of the W-B were significantly related to grade-point average in school. The correlation for Form I was .29, while that for Form II was .26. These *r*'s are based upon a narrow range of talent. Although Merrill and Heathers (**86**) also sampled intelligence narrowly they obtained a correlation of .58 between W-B *V* IQ and grade-point average.

Summary. As time passes the W-B and its later forms seem to attain an increasingly venerable position. It is now the well-accepted standard of intelligence evaluation for adults. As many tests of varied composition are correlated with the W-B it becomes clear that the W-B occupies a central position in evaluating the factor of

general intelligence. When corrections are made for the ever-present restricted ranges of talent, the correlations between the W-B and other tests argue well for both the reliability and validity of the test.

The principles of test construction employed with the W-B have been so successful that an alternate form (**132**) and a children's scale (**133**) have been developed. These two forms of the test seem to be comparable to the W-B in their capacity to evaluate general intelligence. However, the extension of W-B-derived subtest pattern interpretations to these tests has not been too well supported.

The authors once again bemoan the failure of investigators to report the range of intelligence of the samples. We are especially distressed that so few of the writers correct their coefficients for range-of-talent. Too often the raw correlation is meaningless as presented in the report. Too often the tests in the file determine the study rather than the nature of the study determining the data to be collected. The sampling of one study (**31**) is exemplary, while still others show much thought behind the manipulation of experimental variables and the statistical analyses (**3**, **31**, **45**, **56**).

Short Forms

Surprisingly, only three papers (**3**, **29**, **53**) propose new short-form combinations of W-B subtests. First of these is Gurvitz's (**53**) thorough sifting of all possible two-subtest combinations. With a large sample the *I-DS* proved best and was confirmed upon cross validation. The second of these proposals is the practical one by Cotzin and Gallagher (**29**). They show that the four-test form *C-S-PA-BD* correlates .94 with the *FS* for a cross-validation group of defectives. Both the *V* IQ and *P* IQ can be estimated from this form, which is felt to be of particular importance when working with mental defectives. Their regression equations provide appropriate raw scale conversions to IQ. Last is the proposal by Alderdice and Butler (**3**), who developed the *I-D-S-OA* combination so that S-B scores could be best predicted for mental defectives.

Previously proposed short forms are evaluated in an additional three papers (**58**, **59**, **75**). McKenzie (**75**) studied the *V-I-S-BD* form proposed by Kriegmen and Hansen. He concludes this form is a good means of estimating the intelligence of mental defectives and that it provides more diagnostic information than similar instruments. In the same year both Herring (**58**) and Hilden (**59**) reported thorough empirical investigations of previously proposed forms. Herring studied the short forms with both mental-hygiene-clinic patients and normals while Hilden used only mental-hygiene-clinic patients. The more extensive study by Herring (**58**) indicates the best two, three, four, and five subtest combinations.

Summary. This review covers only one third as many studies dealing with short forms as appeared in the previous review (**97**). Instead of a deluge of brief unrelated studies it is gratifying to encounter well-planned comprehensive studies (**3**, **53**, **58**, **59**). Two of the reports (**58**, **59**) involved considerable effort directed at bringing about an integration of work in the area of short forms. One of the proposed short forms (**29**) is the first to yield an estimate of both *V* IQ and *P* IQ for use with defectives. Sampling problems such as range of talent and cross validation have also received more consideration than for-

merly. This greatly increased sophistication is most encouraging.

Applications With Special Population

Sex differences. A comprehensive study by Norman (**89**) explores sex differences in young superiors. Many significant differences in subtests as well as for the *V* and *P* IQ's were encountered. In general, men have higher scores than women, and the greatest difference is apparent on *A*. This relatively large sex discrepancy in *A* was encountered in college students by Guertin (**51**) and proved puzzling. Vane and Eisen (**125**) found that both delinquent and normal girls had lower subscores on *A* and *D* than the standardization population of Wechsler. Strange and Palmer (**120**) employed psychiatric clinic patients for their study of sex differences. Their results agree with Norman's positive findings as do the more recent conclusions of Goolishian and Foster (**48**) using neuropsychiatric patients. Brown and Bryan (**18**) differ in that their normals do not show the superiority of the males so clearly, and patterning is somewhat different than found by other authors.

Miscellaneous populations. Geriatric changes in W-B performance were studied by Birren (**13**) in an unusual way. Instead of merely evaluating subtest scores he studied the factor-scores of geriatric subjects. His Verbal Comprehension Factor held up well with advancing age, while his Closure Factor (performance) showed sizable decrements. As a final comment Birren states, "It is apparent that the W-B does not include tests of many of the known primary abilities. For this reason, it is of limited use in attempting to describe age changes in the intellect."

Colored and white neurotics were matched for the usual variables by Davidson *et al.* (**30**). The *A* subtest and performance scores were lower for the colored. Goldstein (**47**) discusses problems encountered in using the W-B in South Africa. A great many changes had to be made to adapt it to this population. Maleci and Montanari (**79**) present a similar report on the adaptation of the test for use in Italy. A report from Calcutta University (**21**) cites the same sort of problems in adapting the test for India. *V* and *PA* required the most alteration. Aeppli-Tanner (**2**) found she could employ a German translation with Swiss adolescents "without extensive alterations." Gross quantitative results compared well with American norms; however, subtest discrepancies were present.

While not too much has been published about the use of the W-B in vocational counseling, a few studies appeared during this review interval. Of particular importance is Patterson's (**93**) guide for counselors. Mean W-B scores are presented for each of 16 broad occupational groups by Simon and Levitt (**108**). Another by Ladd (**70**) discloses higher *V* IQ's for students in academic teacher-preparation areas as compared with nonacademic teacher-preparation areas. Those in the nonacademic areas showed higher *P* IQ's than those in the academic areas. The authors cautiously suggest there might be some implications for counseling. Merrill and Heathers (**84**) present a normative study setting forth W-B centile scores for a college counseling-center group and also for college student volunteers. They found little difference in the IQ's for the two samples. They concluded that those with personal difficulties were not likely to be significantly lower on the W-B.

Summary. Sex differences with normals and psychiatric subjects

seem to be rather well established. A study employing a representative sample of the general population to identify these sex differences appears much needed at this time. It would be of interest to see whether sex differences appeared in the standardization results of the new Wechsler Adult Intelligence Scale. Perhaps differences in performance by the sexes at one level of intelligence are reversed in direction at a different level so that the total mean differences in test performance are minimal. Differences in *A*, *D*, and *PC* are repeatedly encountered (**48**, **89**, **120**). Researchers and clinicians should take heed of this and specify sex as well as age when proposing patterns or applying pattern analyses.

Some of the more familiar populations of past research have received minimal attention recently, while some new interests appear (**2**, **47**, **49**, **70**, **79**, **84**). It seems appropriate that as the problems of the past are resolved, new interests should appear. Accompanying this progress is the extension of the W-B to foreign populations.

Refinements and Critiques

Administration and scoring. Kitzinger and Blumberg (**66**) present a supplementary guide for administering and scoring the W-B. It includes scoring principles and examples as well as good descriptions of conventional subtest rationale. Cohen (**23**) was interested in scoring and administration biases. He studied samples of W-B's from 13 examiners and was able to identify evidence of examiner biases that would serve to reduce reliability. Those subtests that show lowest reliability were those that produced the most systematic interexaminer disagreement. This raises the possibility that much of the variability of subtests that is ascribed to unreliability may be systematic variance related to the nature of the examiner.

Item order of the various subtests again receives some consideration by Norman (**89**). An incidental finding of his larger study was that the order of difficulty for superiors agreed better with the order found by Jastak than with Wechsler's (**131**). Mech (**83**) presents the order of difficulty for high school students. He suggests that the item order be changed and then appends the statistically wise but impractical condition that this " . . . applies only to students in the 12th grade." Mech is faced with the enigma of establishing an order of difficulty that will apply for any group of subjects. It would seem necessary to revert to Wechsler's intent of establishing the item order on a representative population. Different orders of difficulty found with small restricted portions of the whole population would not then be a basis for challenging the originally proposed order. Russell (**102**) proposes a revised sequence of *V* words for use with neuropsychiatric patients.

Other subtests have been studied. For example, Luchins and Luchins (**74**) examined the effect of varying the instructions on the *DS* subtest. This well-designed study permits the following conclusions: (*a*) initial test score cannot serve as a reliable index of learning, (*b*) speed emphasis may retard performance, and (*c*) Wechsler's directions are so ambiguous that they produce greater variability among the subjects as compared with the experimental directions. Guertin (**51**) was interested in a different subtest. He studied the effect of instructions and item order upon *A*. Contrary to the investigator's expectancy, the superior subjects of the

study were not threatened by encountering difficult items early in the presentation. In fact, they accepted the challenge and performed better than with conventional order of presentation. It seems possible that mentally subnormal subjects might show just the opposite reaction, but this has yet to be verified.

Eglash (**38**) studied the relationship between the *FS* score and the *shoes* item on *C*. He finds some justification for relating the number of reasons given to intelligence. However, the spread between one and two reasons is such that it would be a more appropriate cutting point than the one in current use. Following up on this suggestion, Armstrong (**8**) rescored the *shoes* item, and the *r* between *C* and the *FS* rose from .42 to .48. The *OA* subtest was the subject of the article by Shannon and Rossi (**106**). They suggest a presentation method for *OA* that was completely described several years earlier by Derner and Aborn and was covered in our last review.

Statistical analysis. An item analysis with high school students was made by Mech (**83**). Item order and discriminative values are presented for each item. Stanley (**115**) discusses what he feels "every good clinician should know." The article deals with what to tell the psychiatrist when the *V* IQ and *P* IQ are both larger or smaller than the *FS* IQ. A rather mathematical description of the fallacy of averaging averages points out that averaging the *V* IQ and *P* IQ yields a meaningless "average" IQ that could not be expected to correspond to the *FS* IQ

Cohen (**24**, **26**) cleverly factor-analyzed the W-B subtests using three types of psychiatric patients. He makes a methodological contribution by demonstrating that the factors underlying success on a test are a function of the sample of people taking the test. The material is presented in detail in a relatively nontechnical fashion in another paper (**25**), and it is highly recommended to clinicians. Cohen identified the three factors: Verbal, Nonverbal, and Distractibility. Similar factors were derived by Birren (**13**) for a group of elderly subjects. His factor names were respectively: Verbal Comprehension, Closure, and Rote Memory. In addition a fourth factor is tentatively identified as Induction. Alderdice and Butler (**3**) factor-analyzed the performance of mental defectives but found only a general and a performance factor. Whiteman and Whiteman (**137**) report on the application of Jastak's predetermined clusters: Reality perception, Psychomotor efficiency, Language polarity, and Motivation. Wheeler (**136**) also reports the results of a factor analysis of the W-B and other instruments. Summarizing, there seems to be a fairly good overlap of factors with a few factors appearing only with certain subjects. Most frequently encountered are a verbal, a performance, and an attention factor.

Miscellaneous. Gurvitz (**54**) reached his limit of professional tolerance and wrote a critical article on the defects in standardization sampling and procedure for the W-B. The criticisms are too many to list and reading of the original article is essential for anyone planning to employ data from the standardization. He also scolds Wechsler for his carelessness in changing subtest instructions between editions of the manual yet not presenting any restandardization. Gurvitz makes a most reasonable plea for more adequate standardization sampling in future work.

The relationship between learning ability and W-B performance was investigated by McLean (**76**). By

studying neuropsychiatric patients with both the W-B and conventional learning tasks, the author was able to reach several conclusions: (*a*) higher *V* IQ is indicative of ability to learn in verbal situations, (*b*) higher *P* IQ is indicative of ability to learn in performance situations, and (*c*) subtest variation is inversely related to learning proficiency. The touch of the learning psychologist is also encountered in the experiment by Burik (**19**). He found that motor test-scores, such as rate-of-tapping, were most closely related to *DS*. There was also good evidence that the incidental learning of the symbols was not related to the *DS* score. He concludes that the *DS* subtest should be regarded more as a test of visuomotor coordination than as a test of new learning.

Hypnosis was used by Kline (**67**) to demonstrate the W-B in action. He hypnotically regressed and progressed a 22-year-old woman. W-B's were obtained in these two states as well as at normal age and in a waking state. Under the varying conditions the IQ remained surprisingly constant although weighted scores were necessarily diminished for both the CA's of eight and 65 years. The changes in subtests with progressed age conformed remarkably well to clinical expectancy. The DQ also is appropriate for that age group. Even item six on *BD*, which is not in the correct order of difficulty, was failed when the subject "believed" she was 65. It would be expected that the "conforming" subject would select for failing the last and more complex-appearing design which employs all the blocks. Not much is proved, but this study makes interesting reading. A thesis by Ficca (**39**) relating autonomic features to W-B performance sounded interesting but was not seen.

A series of studies by Stacey and others (**113**, **114**, **115**) investigated Gerstein's hypothesis about conceptual performance. Gerstein designates the descriptive definition as lowest in intellectual requirements, while the functional is intermediate; and the conceptual definition is most associated with high intelligence. Subnormals failed to confirm the predictions of frequency of occurrence for the three types of definitions for *V* (**114**). A similar result was obtained when concentrating on *S* answers (**113**). In the most recent study, Stacey and Spanier (**115**) switched to superiors for a study of *V* responses. The functional was actually associated with lower intelligence than was the descriptive-type definition. This finding was in line with the earlier two studies (**113**, **114**) and suggests that the descriptive-type definition is of a slightly higher level than the functional definition.

Summary. There seems to be a welcome increase in comprehensive, theoretical articles this time (**8**, **13**, **23**, **54**, **76**, **83**, **89**). The sophistication of researchers seems to be advancing. Some evidence of this is found in the relative large number of factor-analytically designed studies in this review (**3**, **13**, **24**, **26**, **136**). Also new is the learning theory design focused upon test theory rationale (**19**, **76**). In addition to the more comprehensive studies, there is an ever-growing body of knowledge about the rationale of the individual subtests.

The Wechsler-Bellevue as a Diagnostic Aid

The use of the W-B as a diagnostic aid implies an assumption most recently expressed by Jastak " . . . intelligence is not a global trait but a general and pervasive part function of the personality." Persistent criticism of such an assumption is ex-

pressed most recently in various research and by Schofield (**105**).

As a Measure of Emotional Factors

Brower in 1947 proposed that a negative correlation existed between the *Hy*, *Hs*, and *Pd* scales of the Minnesota Multiphasic Personality Inventory and the W-B IQ, the probable rationale being that neurotics and persons with character disorders would show intellectual impairment. Winfield (**138**) checked this hypothesis and found no support for it.

Other interest in the relationship between personality tests and intellectual performance is confined to the Rorschach test. Probably most comprehensive is the study by Holzberg and Belmont (**61**). They hypothesized a number of relationships between Rorschach signs and W-B test performance, but only four of their many predictions were substantiated. They set forth their empirical findings under eight features of Rorschach performance with counterparts in W-B performance. A single hypothesis was tested by Burnham (**20**). He failed to find support for the hypothesis that there is a relationship between $H\%$ and *PA*. However, the thesis by Spaner (**111**) touched upon related material. He was able to demonstrate a relationship between *M*'s on the Rorschach and *PA* score. Five of the ten hypothesized relationships between Rorschach and W-B performances were confirmed. Another thesis (**122**) not seen, also deals with the relationship between Rorschach features and W-B performance.

Summary. Attempts to validate the general assumption that something other than a mere IQ may be extracted from performance on the W-B have been disappointing. However, as in the case of the comparison between the Rorschach and the W-B, it might be reiterated that (*a*) the Rorschach scores, as statistically valid measures of personality have not been established, and (*b*) the rationales for the subtests as proposed by Wechsler have been challenged.

As a Measure of Change Following Therapy

Markwell, Wheeler, and Kitzinger (**81**) recorded the W-B performance of schizophrenics before and after prefrontal lobotomy; and Smykal and Wilson (**110**) did the same with electroconvulsive therapy. Markwell, Wheeler, and Kitzinger found no statistically significant differences between pre- and postoperative subtest performance; Smykal and Wilson found some gross group differentiation but considerable individual differences to attenuate the positive aspect of their findings.

In the Smykal study, the W-B was administered before treatment and after the fifth and tenth administration of shock (shock being given twice a week for five weeks). Interesting results demonstrated (*a*) the percentage of IQ's of 35 or below decreased 50 per cent after the fifth shock; the Markwell group found that lobotomy increased the testability of the patients by 20 per cent, (*b*) the highest point of efficiency was at the midpoint of treatment, after the fifth shock; and (*c*) the preshock pattern corresponded to Rapaport's chronic schizophrenic group, while the post-fifth shock collated highly with his acute group.

Summary. Research failed to demonstrate differences in subtest performance reliably following either prefrontal lobotomy or ECT. However, the generalization of the two researches, reviewed in this section, must be delimited by the restricted size and nature of their sample popu-

lations. What the research does attest to is the fact that, as the result of therapy, the patient does become more amenable to testing.

Mental Deterioration Index

Validity. Tests of the validity of Wechsler's (**131**) Mental Deterioration Index (MDI) have produced generally negative results even though Howell (**62**) found that, " . . . it may be tentatively concluded that Wechsler's assumption of deterioration at all levels is correct." Corsini and Fassett (**28**) culled the test protocols for 100 inmates of San Quentin prison. In this cross-sectional study, the total population was divided into twelve five-year periods, starting with ages 15–19 and going to 70–74. These investigators found that by and large the verbal test scores increase with age while performance scores decline, suggesting a *V–P* ratio rather than the "hold," "don't hold" ratio, as the more sensitive measure of mental deterioration. These findings are in good agreement with those reported by Birren (**13**).

Bensberg and Sloan (**10**) tested the validity of Wechsler's MDI on a group of mental defectives. They culled their protocols from the files of the Lincoln State School and Colony, excluding those individuals who had not received S-B IQ's of 42 or better, and who showed evidence of psychotic or organic processes. The results with the W-B were compared with those on the Arthur Point Scale and the S-B. The results suggested (*a*) although the Arthur and the S-B demonstrated a negatively accelerating function with regard to CA increments, the W-B scores tended to increase, rather than decrease, and (*b*) the W-B scores did not differentiate the older group (30–55) from their earlier MDI scores obtained between the ages of 15 and 24. The authors conclude that, " . . . the 'normal deterioration' which Wechsler found for older subjects may have been an artifact due to faulty sampling, at least for the lower intelligence groups."

The study by Fox and Birren (**41**), however, suggests that when psychotic or neurological dysfunctioning is excluded, a group of sixty-year-olds are significantly different in subtest patterning from a group of 69-year-olds. Fox and Birren find that only two of Wechsler's "hold" group "hold," viz., *V* and *C*.

Glik (**46**) offers a very interesting thesis to explain the inability of measures of deterioration, as Wechsler's MDI, to reflect intellectual deterioration adequately. He assumes that the manner of measuring present functioning is in error. Glik reasons that the ability to *recognize* the meaning of items one once knew is a more sensitive measure of deterioration than measuring what items the subject can actually *recall* in the present.

Testing this assumption on the *V* and *I* subtests, Glik found a significant *t* between recall and recognition on *I* items only. Glik generalizes his results to verbal questions without any really adequate substantiation of this point, although his proposals surely merit further study. At any rate, Glik's results do suggest that a discrepancy score between recall and recognition may be more meaningful than the method of recall now used with *V* and *I* items.

In senile psychosis. The major question to be asked here is: does the addition of a psychotic process superimposed upon the "normal"senescent decline significantly alter the W-B performance?

Berkowitz (**11**), Botwinick and Birren (**15**), and Doerken and Greenbloom (**36**), tested the efficiency of the W-B to differentiate "normal" and "abnormal" seniles and found

positive results, while Botwinick and Birren (16) found that the MDI does not differentiate a control group of seniles from an experimental group of seniles with psychosis. The investigators of the latter study as well as Birren (13) emphasize that mental illness in these aged caused a greater deficit in *P* IQ than *V* IQ although there was good evidence of inroads on Verbal performance also.

In organic brain disease. Anderson (6) and Ptacek and Young (94) conducted research in attempts to determine how effective the MDI is in identifying mental deterioration in organics. The results point to a rejection of the sensitivity of the MDI.

Anderson (6) addressed himself to the effects of laterality of localization of brain damage and its effect upon the MDI, with the results being negative. Although the subjects were identified neurologically as organic, they were not cross identified by the MDI

Wheeler and Wilkins (135) tested the Hewson ratio, an empirical formula devised to differentiate organics from nonorganics, but found it to be lacking in ability to make individual predictions.

In neurosis. O'Connor (90) tested the effect of neurosis upon the MDI. The results indicate that the "hold" versus "don't hold" ratio is inapplicable to neurosis.

Summary. Research attempting empirically to define the utility of the MDI has eventuated in negative results. The MDI has proven to be insensitive to deterioration in senility, with or without psychosis, organic brain disease, and neurosis. The research findings did suggest that the present "hold"–"don't hold" ratio was not sensitive to changes due to a deteriorative process, but that the verbal-performance dichotomy was.

Regardless of the nature of the findings per se, as we look back over the various investigations conducted in this particular area with the W-B, one is struck by the number of important methodological criticisms that may be leveled against the experimentation. In the first place, two validity studies use highly restricted sample populations. Both these investigations rejected the utility of the MDI with their populations, but it is hard to generalize as to the validity of the MDI with other populations.

It is most difficult to propose seriously an index of deterioration that would cover the full range of intelligence and age. Probably Wechsler was too ambitious. It is still more difficult to make cross comparisons of studies with such wide ranges of variables which are known to relate to subtest patterning, viz., IQ, CA, education, and vocational skills. Such variables only serve to confound the data and confuse the research worker. Even more difficult to interpret are results from a single sample composed of both psychotics and nonpsychotics, organicity due to trauma, and organicity due to endogenous causes, and even varying degrees of organicity and central involvement (94). Anderson in another study (7), attempting to define the effect of laterality of localization of brain damage, has this to say about his own sample, " . . . nondominant hemisphere sub-group . . . was significantly older than the dominant hemisphere group," and " . . . although brain damage was established beyond any reasonable doubt . . . the criterion of unilaterality is probably relatively poor."

It appears that the conclusions drawn from the results of the investigations attempting to define the effectiveness of the MDI must be tentative because of restricted or questionable sampling. These have apparently introduced distortion into

the subtest performance, confounding the results. Yet there is uniform agreement that the MDI was not effective with those samples employed.

Scatter and Pattern Analysis

Considerable research has been conducted on patterns of performance on the subtests, and on scatter. Since the last review, emphasis has been placed on empirically defined, objectively derived analyses of patterns, and considerable clinical and research acumen has been canalized into this intriguing question.

General findings. Jastak (64) finds that pattern analysis is more reliable if raw, rather than weighted scores, are utilized. Alimena (4) and Jackson (63) use a *z*-score transformation of the weighted scores to alleviate the scatter inherent in the test itself. Bradway and Benson (17) are concerned with the extreme individual deviations present in Rapaport's findings for the diagnostic groups. Jastak (64), Monroe (88), and Wittenborn and Holzberg (139) all found that emotional adjustment was inversely related to the amount of scatter, all using various criterion measures. Collins (27) found a direct relationship between variation on EEG patterns and variation in IQ scores. Purcell, Drevdahl, and Purcell (95) found a Pearson *r* of .31 (significant at .01 level) between measures of anxiety (Hypochondriasis, Depression, and Psychasthenia Scales on the MMPI) and scatter on the subtests, and Moldawsky and Moldawsky (87) found that *D* is an apt indicator of anxiety.

However, neither Shoben (107) nor Matarazzo (82) found any relationship between anxiety and performance on the subtests. Kaldegg (65), Love (73), and Wittenborn and Holzberg (139) found no significant relationship between variability of subtest performance and pathology; and Rakusin (98) found " . . . a lack of uniformity in normal scatter patterns on vocabulary scatter." Rakusin, incidentally, found a significant difference on the patterns of subtests, for a clinic (maladjusted) and nonclinic (adjusted) group, but also discovered that these apparent differences disappeared when the effects of age and IQ on total scatter were alleviated by the method of multiple covariance. Rakusin concludes (98), "The differences obtained could not be attributed to maladjustment." Merrill and Heathers (85) conclude, on the basis of their research on college students, that "scatter is to be expected in groups of above average or superior adults." Jastak found that neurotics, schizophrenics, and organics were not differentiated by Wechsler's signs, and Wittenborn and Holzberg (139), in a chi-square analysis of the performance of paranoid schizophrenics, manic-depressives, alcoholic psychotics, and psychopathic personalities, found their individual group performances indistinguishable.

Holzberg, Alessi, and Talkoff (60) tested the ability of seven judges to predict premorbid intelligence of ten psychotic patients at Connecticut State Hospital. In general, the judges made their evaluations on the basis of the amount of intertest scatter in the individual protocols. Intercorrelations between the judge's estimates ranged from .15 to .87. However, such a matrix of correlations tells little about the over-all intercorrelation of the judges, i.e., how well were they doing as a group. Hence, one of the present writers (GF) simply reproduced the complete correlation matrix from that presented in the article, converted the rho's into ranks, and computed a *W*,

a coefficient of concordance, a non-parametric multiple-rank correlation technique, which yields an over-all coefficient of correlation. *W* was .27 (not significant) hence indicating that there was poor agreement between the judges. The poor results of the Connecticut research may be a function of the inexperience of the judges since four of the seven were interns, the other three being staff members.

Methodology. Before we discuss the research on pattern and scatter analysis with regard to the various disease entities, the writers deem it necessary to preface the discussion. Antecedent to an understanding of the effect of psychopathology upon W-B performance seems to be the need to recognize that factors other than psychopathology affect subtest performance, and must be taken into account in such analysis. For instance, Cohen (**23**) found an examiner bias; Collins (**27**) found variation in subtest as a function of age; many (**18, 48, 51, 89, 109, 120, 125**) found significant deviations on subtest scores due mainly to sex; Aronov (**9**) and O'Connor (**90**) testify to the effect of education on scatter; and Aronov (**9**), French and Hunt (**44**), Merrill and Heathers (**85**), Norman (**89**), O'Connor (**90**), Ortar (**92**), and Schnadt (**104**) validate the effect of IQ level per se on scatter.

The reader will understand why knowledge of these findings is propaedeutic to an adequate purview of the research on pattern analysis with the W-B. In attempts to control for the effects of these variables, the research worker has relied upon the statistical concept of randomization. Statistically the assumption is maintained that if you randomize the effects of these uncontrollable, hence, unmeasurable parameters, equally throughout the cells of one's table, the effects of the randomization will be to equalize the effect of these parameters upon the variance. Such statistical reasoning has been translated into research methodology in the form of the matched group design. However, certain research suggests that the simple randomized design is not the answer to a research worker's prayers.

Frank, Corrie, and Fogel (**43**) and Reich (**99**) have demonstrated that we are no longer meaningfully speaking about "neurotics" or "schizophrenics" when we include wide ranges of attributes (age, education, IQ) in one group. For instance, Reich (**99**) demonstrated that when he compared the subtest performance of his group of schizophrenics from Kings County Hospital, with data from other groups of subjects, similarly diagnosed, presented in previous published literature, the rank-order correlations ranged from .84 to .19. Frank, Corrie, and Fogel (**43**) performed an analysis of variance on the subtest performance of like-diagnosed cases presented in the literature which included wide ranges of IQ, education, age, etc., and found them to be statistically different. If one begins by comparing incomparables, the results are bound to be spurious.

Subtest distortion is not just a function of research methodology, but may accrue as a result of factors inherent in the W-B itself. For instance, Cohen (**26**) finds that the W-B subtests do not always appear to measure the same factors in different types of patients. Marks (**80**) criticized Wechsler's qualitative pattern approach to scatter analysis, since it permits of a wide range of subtest variation coding within a given diagnostic category, such as from $++$ to $-$. This tends to "load the data" in terms of a lack of reliability and validity of Wechsler's diagnostic pattern. It appears "safe" to review re-

search on pattern analysis in the different diagnostic categories now that the "warning alarm" has been sounded.

Schizophrenic patterning. Harper (**55**) applied Fisher's discriminant function to pattern analysis. In general, Harper felt that the pattern of subtests differentiated his group of schizophrenics from his normals, and further, that the subtypes, paranoid, hebephrenic, and catatonic, were significantly differentiated from the rest of the schizophrenics and the normals.

Harper wrote, as regards the regression equation he had formulated, "The extent to which the regression weights would be applicable to a new sample from a different hospital . . . is not known. . . ." Cross validation was needed, so on suggestion by Solomon Machover, Reich (**99**) performed this test. He found that the Harper equation identified 65 per cent of a new schizophrenic population at Kings County Hospital as schizophrenic, concluding that, in general, individual differentiation was poor. On a further suggestion by Machover, Frank (**42**) tested the effectiveness of Harper's equation when applied to a heterogeneous group of psychotics (excluding schizophrenics) at Kings County Hospital. He found that the Harper equation misidentified 47 per cent of this population as schizophrenic, thereby questioning the utility of the formula.

Rogers (**100**) compared schizophrenics and neurotics on fifteen signs postulated by Rabin, Rapaport, and Wechsler, and found only eight of them tenable. However, Rogers grouped the various subtypes of schizophrenia under a single heading. On the other hand, on the basis of a comparison of 50 neurotics and 50 schizophrenics, little agreement was found between Wechsler pattern and psychiatric diagnosis. Whiteman and Whiteman (**137**) had some success using factor analysis on the performance of 50 schizophrenics and 50 police applicants, finding that clusters entitled "reality perception" and "psychomotor efficiency" significantly differentiated the groups. McNeal (**77**) checked Wechsler's signs for schizophrenia on equated normal and schizophrenic groups for 340 male veterans of World War II. "Only one of Wechsler's signs discriminates significantly between the . . . groups . . . but it identifies schizophrenics more normal than normals." What appears to be a very interesting and thought-provoking finding is Monroe's (**88**) demonstration that schizophrenics with average or above average intellectual level had no more scatter than neurotics, and that extreme scatter was a characteristic of schizophrenics only with low obtained IQ.

Patterning in the affective disorders. The writers could find but one investigation in this area during the entire five-year period, an apparent sign of the need for more research. Waldfogel and Guy (**126**) found differences between manic-depressives and controls, but none between depressive or manic states of the manic-depressive psychosis. However, further analysis revealed that the parameter of age was contributing an unknown amount of variance to the pattern, hence depreciating the meaningfulness of the results.

Patterning in neurosis. Monroe (**88**) using a factorial design to test the effect of levels of adjustment, intelligence, and locality, found little differentiation between neurotics and well-adjusted normals when the interaction of these effects were deleted from the between-group variance. Schillo (**103**) found no difference between neurotics and normals on dis-

parity between verbal and performance IQ, performance on individual tests, or rank-order correlation of test performance, although his neurotics showed greater variability from its own group mean than the normals. Schillo's subjects, though, included individuals subject to anxiety reaction, obsessive-compulsive reaction, and mixed neurosis; and no data are offered as to the possible differences within this group as related to the subtypes.

Patterning in mental deficiency. Alderdice and Butler (**3**) seem to be the only investigators interested in identifying a mental defective pattern. They present thoughtful approaches to pattern analysis and compare their findings with those of previous workers. Their pattern for mental deficiency, while not a sharply delineated one, is similar to those previously reported. It is most at variance with the pattern initially proposed by Wechsler (**131**). Little work seems to have been done in terms of an understanding of the performance of the mentally defective subject on the subtests of the W-B. One investigation, by McPherson and Fisch (**78**) attempts to understand the defective's poor performance on *S*. Using approximately 30 subjects with an IQ range of from 55 to 85, they found that 66 per cent of the failures were due to evasiveness, what McPherson and Fisch term "learned negativism." The authors suggest that as a result of such behavior, testing may not yield an exact measure of the defective's ability.

Patterning in organicity. Collins (**27**) investigating the effect of epileptic involvement found that only cases diagnosed with organicity demonstrated deterioration. Further analysis revealed the superiority of the performance of the endogenous to the exogenous group. A comparison of Collins' 400 outpatient epileptics with the protocols of institutionalized groups of epileptics, psychotics, neurotics, and psychopaths, presented in previous research, showed considerable differences. In the analysis of 26 cerebral arteriosclerotic patients compared with 26 patients with "other forms of cerebral pathology" matched for age, education, and IQ, Oppenheim (**91**) found no significant differences on subtests or in total variability of performance.

In one investigation Diers and Brown (**35**) found that there was an inverse relationship between W-B IQ and the Hughes factor-analytically derived signs on the Rorschach for intracranial pathology. In another investigation Diers and Brown (**34**) analyzed the protocols of 24 patients diagnosed as having multiple sclerosis. When compared with normal controls, the sclerotic group demonstrated a lower memory span for *D* and superior *PC;* however, a Pearson *r* of .17 (not significant) was obtained between Wechsler's organic signs and actual organicity. The authors offer the following alternatives to explain their results: "Quantitative signs of organic damage to the brain, or the index of deterioration on the Wechsler-Bellevue scale, are inadequate as an indicator of existing cortical damage in multiple sclerosis," or "No cortical pathologic changes existed in the population with multiple sclerosis comprising this study."

Patterning of the sociopath. The patterning of the sociopath (psychopath, delinquent) has come under considerable scrutiny since the last review. Bernstein and Corsini (**12**) tested the validity of Wechsler's assumption that performance scores are higher than verbal scores—which they could not reject. By the method of forward and backward presentation of the subtests, they rejected

the assumption that this was a spurious difference due only to the fact that the performance part of the test is administered late in the examination period permitting the sociopath to adjust more adequately.

Graham (**49**) found that there was a significant relationship between performance on the W-B and school performance, and that the scattergram of Wechsler's adolescent psychopath closely approximates that of the unsuccessful reader. Graham makes the appealing suggestion, "It does not seem unreasonable to assume that the profile (Wechsler's adolescent psychopath) is typical of educationally retarded youth without regard to his moral qualities."

Vane and Eisen (**125**) tested the difference on the W-B between matched groups (age, intelligence, socioeconomic background) of delinquent and nondelinquent girls, and found none. Higher scores on *D* and *I* for the nondelinquent than for the delinquent group was about all they found to differentiate the two groups.

Gurvitz (**52**) compared the subtest performance of a matched group of inmates diagnosed as psychopaths and "nonpsychopathic" inmates, and Wechsler's standardization group of psychopaths. Through a chi-square and *t*-test of significance, the results indicated that there was no characteristic subtest pattern for the psychopath, and, further, that there were no differences between the "psychopathic" and the "normal" prison inmate, and between these two and Wechsler's psychopathic population.

In a similar study, Clark and Moore (**22**) had previously attempted to differentiate the test patterns of subclassifications of military offenders (no NP disorder, immaturity reaction, and pathological personality types—presumably psychopaths) but with no success.

If one considers unauthorized discharge from a medical hospital as asocial or antisocial behavior, Thurston and Claden's (**123**) study on the irregular discharge of tuberculosis patients becomes pertinent here. However, the W-B could not differentiate the "AWOL" group from those who remained in the hospital, either in terms of IQ or subtest patterning.

Summary. Our somewhat jaundiced eye continues to reject the assumption of unique subtest performances by schizophrenics. The results are at best inconsistent, but their very inconsistency may testify to the erroneous assumption that randomization will control for the effect of the parameters upon subtest performance. But the reader of such researches is left in the predicament of being uncertain whether to attribute the negative results to either the methodology or to the hypothesis.

For groups of mental defectives patterning seems better established than for schizophrenics. The search for patterns of performance characteristic of the organic brain-damaged, the manic depressives, and neurotics have been rather fruitless but, as with schizophrenics, the sampling methods have left much to be desired.

Researches with the sociopath indicate some points of agreement but there is still much disagreement. The pattern proposed by Wechsler, "adolescent psychopath," is coming to be evaluated in terms of the subjects' academic training, motivation, and general background of experience.

The over-all impression gained from the pattern studies reviewed is that the findings are inconclusive. Patterns suitable for clinical use will not be forthcoming until methodological improvements appear. Further advances in typology can do much to establish suitable diagnostic

criteria for which patterns can legitimately be sought.

General Summary

The past five years exhibit two major changes in the research trends with the W-B. In the first place, there has been a realignment of the general nature of the studies. Research with psychiatric syndromes has been reduced, whereas a larger proportion of the studies reviewed deal with the W-B as a test of general intelligence, investigating its reliability, validity, rationale, etc. Secondly, it has been noted throughout the review that the number of well-controlled and statistically sophisticated studies has markedly increased.

One rather new and somewhat unexpected trend is the demonstration of sex differences. It points up an additional uncontrolled factor in previous studies and has important implications for the use of test patterns, scatter, and such. Of course, the need for a larger, more representative sample for the demonstration of sex differences, or their absence, at different levels of intelligence still remains. It may be added that another factor that needs control in research is that of socioeconomic level. Social class has been found to have a bearing on intelligence test achievement. This remains comparatively unexplored with the W-B. Moreover, it may have an important relationship to the Verbal-Performance discrepancies noted in psychopaths, delinquents, and others whose status in such classifications may be related to socioeconomic level as well.

When one looks at the work done with various psychiatric populations by means of scatter and patterning methods, one can readily conclude that "nothing new has been added" either in methodology or in definite findings. The results are still "inconclusive." It may be wondered whether it should not be said that there are no positive results instead of leaving an open crack in the door, implicitly indicating a still-tenacious clinging to an overworked hypothesis.

Maybe, with the creation of a newly standardized instrument, similar in structure to the W-B, but not suffering from the numerous weaknesses which have been pointed up in this review and in previous ones, more fruitful research with pattern analysis will be forthcoming. However, such research cannot continue to close the eyes to the weakness of the criterion itself—psychiatric diagnostic classification.

BIBLIOGRAPHY

1. Abrams, E. N. Prediction of intelligence from certain Rorschach factors. *J. clin. Psychol.*, 1955, **11**, 81–83.
2. Aeppli-Tanner, Louise. Anwendung des Wechsler-Bellevue tests in der Schweiz [(Use of the W-B in Switzerland)]. *Schweiz. z. Psychol. Anwend.*, 1954, **13**, 136–145.
3. Alderdice, E. T., & Butler, A. J. An analysis of the performance of mental defectives on the Revised Stanford-Binet and the W-B Intelligence Scale. *Amer. J. ment. Def.*, 1952, **56**, 609–614.
4. Alimena, B. Norms for scatter analysis on the Wechsler Intelligence Scales. *J. clin. Psychol.*, 1951, **7**, 289–290.
5. Allen, R. M., Thornton, T. E., & Stenger, C. A. Ammons and Wechsler test performances of college and psychiatric subjects. *J. clin. Psychol.*, 1954, **10**, 378–379.
6. Anderson, A. L. The effect of laterality localization of brain damage on W-B indices of deterioration. *J. clin. Psychol.*, 1950, **6**, 191–194.
7. Anderson, A. L. The effect of laterality localization of focal brain lesions on the W-B subtests. *J. clin. Psychol.*, 1951, **7**, 157–162.

8. ARMSTRONG, RENATE G. A comparison of the Comprehension subtests of the W-B Intelligence Scale, Forms I and II. *J. clin. Psychol.*, 1953, **9**, 172–176.
9. ARONOV, B. Education and intelligence variables in the psychopath's W-B Intelligence Test achievement. Paper read at Midwest. Psychol. Ass., Cleveland, April 1952.
10. BENSBERG, G. J., & SLOAN, W. A study of Wechsler's concept of "normal deterioration" in older mental defectives. *J. clin. Psychol.*, 1950, **6**, 359–362.
11. BERKOWITZ, B. The W-B performance of white males past age 50. *J. Geront.*, 1953, **8**, 76–80.
12. BERNSTEIN, R., & CORSINI, R. J. W-B Patterns of female delinquents. *J. clin. Psychol.*, 1953, **9**, 176–179.
13. BIRREN, J. E. A factorial analysis of the W-B Scale given to an elderly population. *J. consult. Psychol.*, 1952, **16**, 399–405.
14. BOTWINICK, J. W-B split-half subtest reliabilities: differences in age and mental status. *J. consult. Psychol.*, 1953, **17**, 225–228.
15. BOTWINICK, J., & BIRREN, J. E. Differential decline in the W-B subtests in the senile psychoses. *J. Geront.*, 1951, **6**, 365–368.
16. BOTWINICK, J., & BIRREN, J. E. The measurement of intellectual decline in the senile psychoses. *J. consult. Psychol.*, 1951, **15**, 145–150.
17. BRADWAY, KATHERINE, & BENSON, S. The application of the method of extreme deviations to Rapaport's Wechsler-Bellevue data. *J. clin. Psychol.*, 1955, **11**, 285–291.
18. BROWN, M. H., & BRYAN, G. ELIZABETH. Sex differences in intelligence. *J. clin. Psychol.*, 1955, **11**, 303–306.
19. BURIK, T. E. Relative roles of the learning and motor factors involved in the digit symbol test. *J. Psychol.*, 1950, **30**, 33–42.
20. BURNHAM, CATHERINE A. A study of the degree of relationship between Rorschach *H*% and W-B picture arrangement scores. *Rorschach Res. Exch.*, 1949, **13**, 206–209.
21. CALCUTTA UNIVERSITY, DEPT. OF PSYCHOLOGY, APPLIED PSYCHOLOGY SECTION. *Report on the working of the section, 1943–1952.* Calcutta: Univer. College of Science, 1953.
22. CLARK, J. H., & MOORE, J. H. The relationship of W-B patterns to psychiatric diagnosis of Army and Air Force prisoners. *J. consult. Psychol.*, 1950, **14**, 493–495.
23. COHEN, E. Is there examiner bias on the W-B? *Proc. Okla. Acad. Sci.*, 1950, **31**, 150–153.
24. COHEN, J. A comparative analysis of the factors underlying intelligence test performance of different neuropsychiatric groups (etc). *Microfilm Abstr.*, 1950, **10**, 313–315.
25. COHEN, J. A factor-analytically based rationale for the W-B. *J. consult. Psychol.*, 1952, **16**, 272–277.
26. COHEN, J. Factors underlying W-B performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 359–365.
27. COLLINS, A. L. Epileptic intelligence, *J. consult. Psychol.*, 1951, **15**, 392–399.
28. CORSINI, R. J., & FASSETT, KATHERINE K. The validity of Wechsler's Mental Deterioration Index. *J. consult. Psychol.*, 1952, **16**, 462–468.
29. COTZIN, M., & GALLAGHER, J. The Southbury scale: a valid abbreviated W-B for mental defectives. *J. consult. Psychol.*, 1950, **14**, 358–364.
30. DAVIDSON, K. S., GIBBY, R. G., MCNEIL, E. B., SEGAL, S. J., & SILVERMAN, H. A preliminary study of Negro and white differences on Form I of the W-B scale. *J. consult. Psychol.*, **14**, 489–491.
31. DELATTRE, LOIS, & COLE, D. A comparison of the WISC and the W-B. *J. consult. Psychol.*, 1952, **16**, 228–230.
32. DELP, H. A. Correlations between the Kent EGY and the Wechsler batteries. *J. clin. Psychol.*, 1953, **9**, 73–75.
33. DESAI, M. M. The relationship of the W-B Verbal Scale and the Progressive Matrices Test. *J. consult. Psychol.*, 1955, **19**, 60.
34. DIERS, W. C., & BROWN, C. C. Psychometric patterns associated with multiple sclerosis. *Arch. Neurol. Psychiat.*, 1950, **63**, 760–767.
35. DIERS, W. C., & BROWN, C. C. Rorschach "organic signs" and intelligence level. *J. consult. Psychol.*, 1951, **15**, 343–345.
36. DOERKEN, H., JR., & GREENBLOOM, GRACE C. Psychological investigation of senile dementia: II. the W-B adult intelligence scale. *Geriatrics*, 1953, **8**, 324–333.
37. DUNCAN, J. O. Correlation between the W-B and Stanford-Binet vocabulary lists. *Med. Tech. Bull.*, 1953, **4**, 45–47.
38. EGLASH, A. Validation of the Wechsler

"shoes" item. *J. abnorm. soc. Psychol.*, 1950, **45**, 733–734.

39. FICCA, S. C. Relationship of "autonomic" blood pressure pattern types of subject's performance on the W-B and the Rorschach Test. *Pennsylvania State College, Abstracts of doctoral dissertations, 1950*, State College, Pa., 1951, **13**, 398–400.
40. FINKELSTEIN, M., GERBOTH, R., & WESTERHOLD, RUTH. Standardization of a short form of the Wechsler Vocabulary subtest. *J. clin. Psychol.*, 1952, **8**, 133–135.
41. FOX, CHARLOTTE, & BIRREN, J. E. The differential decline of subtest scores of the W-B Intelligence Scale in 60–69-year-old individuals. *J. genet. Psychol.*, 1950, **77**, 313–317.
42. FRANK, G. H. Patterning of the schizophrenic on the W-B intelligence test. *Psychol. Newsltr*, 1953, No. **46**, 7–12.
43. FRANK, G. H., CORRIE, C. C., & FOGEL, J. An empirical critique of research with the W-B in differential psychodiagnosis. *J. clin. Psychol.*, 1955, **11**, 291–293.
44. FRENCH, ELIZABETH G., & HUNT, W. A. The relationship of scatter in test performance to intelligence level. *J. clin. Psychol.*, 1951, **7**, 95–98.
45. GERBOTH, RENATE. A study of two forms of the W-B Intelligence Scale. *J. consult. Psychol.*, 1950, **14**, 365–370.
46. GLIK, E. E. A comparison of recall and recognition types of measurement on verbal items, and their implications for deterioration testing. *J. clin. Psychol.*, 1951, **7**, 157–162.
47. GOLDSTEIN, M. J. A preliminary evaluation of the use of the W-B Adult Intelligence Scale in South Africa. *J. soc. Res., Pretoria*, 1950, **1**, 220–226.
48. GOOLISHIAN, H. A., & FOSTER, A. A note on sex differences on the W-B Test. *J. clin. Psychol.*, 1954, **10**, 298–299.
49. GRAHAM, E. E. W-B and WISC scattergrams of unsuccessful readers. *J. consult. Psychol.*, 1952, **16**, 268–271.
50. GRZYWAK-KACZYNSKA, M. Metoda Wechsler-Bellevue; ako pomoc w diagnozie psychologicznej [W-B method as a help in psychological diagnosis.] *Zdrowie psych.*, 1950, **4** (3–4), 53–76.
51. GUERTIN, W. H. The effect of instructions and item order on the W-B arithmetic subtest. *J. genet. Psychol.*, 1954, **85**, 79–83.
52. GURVITZ, M. S. The W-B test and the diagnosis of psychopathic personality. *J. clin. Psychol.*, 1950, **6**, 397–401.
53. GURVITZ, M. S. The Hillside Short Form of the W-B. *J. clin. Psychol.*, 1951, **7**, 131–134.
54. GURVITZ, M. S. Some defects of the W-B. *J. consult. Psychol.*, 1952, **16**, 124–126.
55. HARPER, A. E. *Differential patterns in schizophrenia on the W-B Intelligence Test.* Allahabad U.P., India: Allahabad Christian Press, Ltd., 1950.
56. HAYS, W., & SCHNEIDER, B. A test-retest evaluation of the Wechsler Forms I and II with mental defectives. *J. clin. Psychol.*, 1951, **7**, 140–143.
57. HELMICK, J. S. Reliability or variability? *J. consult. Psychol.*, 1952, **16**, 154–155.
58. HERRING, F. H. An evaluation of published short forms of the W-B scale. *J. consult. Psychol.*, 1952, **16**, 119–123.
59. HILDEN, A. H., TAYLOR, J. W., & DUBOIS, P. H. Empirical evaluation of short W-B scales. *J. clin. Psychol.*, 1952, **8**, 323–331.
60. HOLZBERG, J. D., ALESSI, S. L., & TALKOFF, A. Judgments of premorbid intellectual functioning in severely impaired psychiatric patients. *J. clin. Psychol.*, 1954, **10**, 219–224.
61. HOLZBERG, J. D., & BELMONT, LILLIAN. The relationship between factors on the W-B and Rorschach having common psychological rationale. *J. consult. Psychol.*, 1952, **16**, 23–29.
62. HOWELL, R. J. Changes in Wechsler subtest scores with age *J. consult. Psychol.*, 1955, **19**, 47–50.
63. JACKSON, C. V. Estimating impairment on W-B subtests. *J. clin. Psychol.*, 1955, **11**, 137–143.
64. JASTAK, J. Ranking Bellevue subtest scores for diagnostic purposes. *J. consult. Psychol.*, 1953, **17**, 403–410.
65. KALDEGG, A. The Wechsler test in clinical practice: comparison of psychiatric and psychosomatic disorders with a control population. *J. ment. Sci.*, 1950, **96**, 908–922.
66. KITZINGER, HELEN, & BLUMBERG, E. Supplementary guide for administering and scoring the W-B Intelligence Scale (Form I). *Psychol. Monogr.*, 1951, **65**, No. 2 (Whole No. 319).
67. KLINE, M. V. Hypnosis and age progression: A case report. *J. genet. Psychol.*, 1951, **78**, 195–205.
68. KNOPF, I. J., MURFETT, BETTY J., & MILSTEIN, V. Relationship between

the W-B Form I and the WISC. *J. clin. Psychol.*, 1954, **10**, 261–263.

69. KNOTT, J. R., CANNICOTT, R. J., UMBERGER, J. P., & BILODEAU, INA McD. Brief tests of intelligence in the psychiatric clinic. *J. clin. Psychol.*, 1951, **7**, 123–126.
70. LADD, A. H. The differential productive value of the W-B Scale for certain areas of teacher preparation. In *Indiana Univer., Sch. Educ., Stud. Educ. 1950*. Bloomington, Ind., 1951, **2**, 62–68.
71. LEVINE, B., & ISCOE, I. A comparison of Raven's Progressive Matrices with a short form of the W-B. *J. consult. Psychol.*, 1954, **18**, 10.
72. LEVINE, B., & ISCOE, I. The Progressive Matrices (1938), the Chicago Non-Verbal and the W-B on an adolescent deaf population. *J. clin. Psychol.*, 1955, **11**, 307–308.
73. LOVE, LENORE R. An analysis of W-B items score for assessing the clinical significance of variability of performance. Unpublished doctor's dissertation, Univer. of California, Los Angeles, 1953.
74. LUCHINS, A. S., & LUCHINS, EDITH H. Effects of varying the administration of the digit symbol subtest of the W-B Intelligence Scale. *J. gen. Psychol.*, 1953, **49**, 125–142.
75. McKENZIE, R. E. A study of the W-B Intelligence Scale and the VIBS Short Form in an institute for the mentally deficient. *Amer. J. ment. Defic.*, 1951, **56**, 174–176.
76. McLEAN, O. S. Divergent scores on the W-B Scale as indicators of learning ability. *J. clin. Psychol.*, 1954, **10**, 264–266.
77. McNEAL, B. F. The prediction of psychiatric diagnoses by signs derived from scatter on the W-B Adult Intelligence Scale. *Dissertation Abstr.*, 1952, **12**, 595.
78. McPHERSON, M. W., & FISCH, R. Affect in the etiology and maintenance of mental deficiency. *J. clin. Psychol.*, 1955, **11**, 55–60.
79. MALECI, O., & MONTANARI, M. Su alcune modificazioni nei punteggi del test Wechsler-Bellevue per il suo impiego nella provincia di Padova [Some modifications in the scoring of W-B for its uses in the province of Padua]. *Arch. Psicol. Neurol. Psichiat.* 1952, **13**, 584–606.
80. MARKS, M. R. A criticism of the W-B Scale as a diagnostic instrument. *J. gen. Psychol.*, 1953, **49**, 143–152.
81. MARKWELL, E. D., JR., WHEELER, W. M., & KITZINGER, HELEN. Changes in W-B test performance following prefrontal lobotomy. *J. consult. Psychol.*, 1953, **17**, 229–231.
82. MATARAZZO, R. G. The relationship of manifest anxiety to W-B subtest performance. *J. consult. Psychol.*, 1955, **19**, 218.
83. MECH, E. Item analysis and discriminative value of selected W-B subtests. *J. educ. Res.*, 1953, **47**, 241–260.
84. MERRILL, R. M., & HEATHERS, LOUISE B. Centile scores for the W-B Intelligence Scale on a university counseling center group. *J. consult. Psychol.*, 1952, **16**, 406–409.
85. MERRILL, R. M., & HEATHERS, LOUISE B. Deviations of Wechsler-Bellevue subtest scores from vocabulary level in university counseling-center clients. *J. consult. Psychol.*, 1952, **16**, 469–472.
86. MERRILL, R. M., & HEATHERS, LOUISE B. A comparison of the Wechsler-Bellevue and ACE tests on a university counseling center group. *J. consult. Psychol.*, 1953, **17**, 63–66.
87. MOLDAWSKY, S., & MOLDAWSKY, P. C. Digit span as an anxiety indicator. *J. consult. Psychol.*, 1952, **16**, 115–118.
88. MONROE, J. J. The effects of emotional adjustment and intelligence upon Bellevue scatter. *J. consult. Psychol.*, 1952, **16**, 110–114.
89. NORMAN, R. D. Sex differences and other aspects of young superior adult performance on the W-B. *J. consult. Psychol.*, 1953, **17**, 411–418.
90. O'CONNOR, J. P. The Wechsler-Bellevue Intelligence Scale as a measure of mental deterioration in psychoneurotics. Unpublished master's thesis, Catholic Univer., 1949.
91. OPPENHEIM, H. Diagnostic limitations of the adult W-B Intelligence Scale in cerebral pathology. *U.N.C. Rec.*, 1952, No. 506, 223–224.
92. ORTAR, G. Hamishmaut hadiagnostit shel mivhan Wechsler etsel baaley ramot sihlixot shonot [The diagnostic value of the W-B test at different intelligence levels]. *M'gamot*, 1952/53, **4**, 199–215.
93. PATTERSON, C. H. The W-B Scales; a guide for counselors. Springfield, Ill.: Charles C Thomas, 1953.
94. PIACEK, J. E., & YOUNG, FLORENCE M. Comparison of the Grassi Block Substitution Test with the W-B in the

diagnosis of organic brain damage. *J. clin. Psychol.*, 1954, **10**, 375–378.

95. PURCELL, C. K., DREVDAHL, J., & PURCELL, K. The relationship between attitude—I.Q. discrepancy and anxiety, *J. clin. Psychol.*, 1952, **8**, 82–85.
96. RABIN, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.
97. RABIN, A. I., & GUERTIN, W. H. Research with the Wechsler-Bellevue Test: 1945–1950. *Psychol. Bull.*, 1951, **48**, 211–248.
98. RAKUSIN, J. The analysis of scatter on the W-B Adult Intelligence Scale in a group of adjusted and a group of maladjusted college students. Unpublished master's thesis, Pennsylvania State Coll., 1949.
99. REICH, H. The applicability of regression equations to the Bellevue Adult Scale in diagnosis of schizophrenia. Unpublished master's thesis, Brooklyn Coll., 1951.
100. ROGERS, L. S. Differences between neurotics and schizophrenics on the W-B Scale. *J. consult. Psychol.*, 1951, **15**, 151–153.
101. RUBIN, H. A quantitative study of the HTP and its relationship to the W-B Scale. *J. clin. Psychol.*, 1954, **10**, 35–38.
102. RUSSELL, G. E. W-B Vocabulary subtest items: revised order of words. *Med. Tech. Bull.*, 1954, **5**, 143–148.
103. SCHILLO, R. J. W-B results of normals and neurotics with obsessive-compulsive features. Unpublished master's thesis, Catholic Univer., 1951.
104. SCHNADT, F. Certain aspects of W-B scatter at low IQ levels. *J. consult. Psychol.*, 1952, **16**, 456–461.
105. SCHOFIELD, W. Critique of scatter and profile analysis of psychometric data. *J. clin. Psychol.*, 1952, **8**, 16–22.
106. SHANNON, W., & ROSSI, P. D. Suggestions for efficient presentation of the W-B Object-Assembly subtest. *J. clin. Psychol.*, 1952, **8**, 413–415.
107. SHOBEN, E. J., JR. The W-B in the detection of anxiety: A test of the Rashkis-Welsh hypothesis. *J. consult. Psychol.*, 1950, **14**, 40–45.
108. SIMON, L., & LEVITT, E. A. The relation between W-B IQ scores and occupational area. *Occupations*, 1950, **29**, 23–25.
109. SMITH, A. E. A comparison of the SRA Primary Mental Abilities Test with the W-B Intelligence Scale. Unpublished master's thesis, Illinois State Normal Univer., 1949.
110. SMYKAL, A., & WILSON, M. O. W-B subtest score changes resulting from electric convulsive therapy. *Proc. Okla. Acad. Sci.*, 1950, **31**, 148–149.
111. SPANER, F. E. An analysis of the relationship between some Rorschach test determinants and subtest scores on the W-B Adult Scale. Unpublished doctor's dissertation, Purdue Univer., 1950.
112. SPITZ, —. De Wechsler-Bellevue Test [The W-B Test]. *Ned. Tijdschr. Psychol.*, 1950, 126–136.
113. STACEY, C. L., & MARKIN, K. A study of the differential responses among three groups of subnormals on the similarities subtest of the Wechsler Intelligence Scale. *Amer. J. ment. Defic.*, 1951, **56**, 424–428.
114. STACEY, C. L., & PORTNOY, B. A study of the differential responses on the Vocabulary sub-test of the W-B Intelligence Scale. *J. clin. Psychol.*, 1951, **7**, 144–148.
115. STACEY, C. L., & SPANIER, S. W. Differential responses among college students on the Vocabulary subtest of the Wechsler Intelligence Scale. *J. educ. Psychol.*, 1954, **45**, 29–35.
116. STANLEY, J. C. Why W-B Full-Scale IQ's are more variable than averages of Verbal and Performance IQ's. *J. consult. Psychol.*, 1953, **17**, 419–420.
117. STEISEL, I. M. The relationship between test and retest scores on the W-B Scale (Form I) for selected college students. *J. genet. Psychol.*, 1951, **79**, 155–162.
118. STEISEL, I. M. Retest changes in W-B scores as a function of the time interval between examinations. *J. genet. Psychol.*, 1951, **79**, 199–203.
119. STORRS, SIBYLL V. Evaluative data on the G.A.T.B. *Personnel Guid. J.*, 1952, **31**, 87–90.
120. STRANGE, F. B., & PALMER, J. O. A note on sex differences on the W-B tests. *J. clin. Psychol.*, 1953, **9**, 85–87.
121. TAMMINEN, A. W. A comparison of the Army General Classification Test and the W-B Intelligence Scales. *Educ. psychol. Measmt*, 1951, **11**, 646–655.
122. TATOM, MARY H. Relationships between W-B sub-test scores and certain Rorschach test factors in clinical patients. Unpublished master's thesis, Catholic Univer., 1949.
123. THURSTON, J. R., & CLADEN, G. Intelli-

gence factors in irregular discharge among tuberculosis patients. *J. consult. Psychol.*, 1954, **18**, 404.

124. VANDERHOST, LEONETTE, SLOAN, W., & BENSBERG, G. J. Performance of mental defectives on the W-B and the WISC. *Amer. J. ment. Defic.*, 1953, **57**, 481–483.
125. VANE, JULIA R., & EISEN, VIRGINIA W. W-B Performance of delinquent and nondelinquent girls. *J. consult. Psychol.*, 1954, **18**, 221–225.
126. WALDFOGEL, S., & GUY, W. W-B subtest scatter in the affective disorders. *J. clin. Psychol.*, 1951, **7**, 135–139.
127. WATSON, R. I. Tests of intelligence. D. W-B Intelligence Scale for adolescents and adults. In A. Weider, *Contributions toward medical psychology*. New York: Ronald, 1953.
128. WEBB, W. B. Corrections for variability: a reply. *J. consult. Psychol.*, 1952, **16**, 156.
129. WEBB, W. B., & DE HAAN, H. W-B split-half reliabilities in normals and schizophrenics. *J. consult. Psychol.*, 1951, **15**, 68–71.
130. WEBB, W. B., & HANER, C. Quantification of the W-B vocabulary sub-test. *Educ. psychol. Measmt*, 1949, **9**, 693–707.
131. WECHSLER, D. *The measurement of adult intelligence* (3rd Ed.). Baltimore: Williams & Wilkins, 1944.
132. WECHSLER, D. *The Wechsler-Bellevue Intelligence Scale, Form II*. New York: Psychological Corp., 1946.
133. WECHSLER, D. *Wechsler Intelligence Scale for Children; manual*. New York: The Psychological Corp., 1949.
134. WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation, 1955.
135. WHEELER, J. L., & WILKINS, W. L. The validity of the Hewson ratios. *J. consult. Psychol.*, 1951, **15**, 163–166.
136. WHEELER, W. M. The internal structure of three clinical instruments. *Amer. Psychologist*, 1950, **5**, 470. (Abstract)
137. WHITEMAN, M., & WHITEMAN, DORIT. The application of cluster analysis to the W-B Scale. *Delaware St. med. J.*, 1949, **21**, 174–176.
138. WINFIELD, D. L. The relationship between IQ scores and Minnesota Multiphasic Personality Inventory Scores. *J. soc. Psychol.*, 1953, **38**, 299–300.
139. WITTENBORN, J. R., & HOLZBERG, J. D. The W-B and descriptive diagnosis. *J. consult. Psychol.*, 1951, **15**, 325–329

Received August 3, 1955

PSYCHOLOGICAL BULLETIN
Vol. 53, No. 3, 1956

THE INTERMEDIARY KEY IN THE ANALYSIS OF INTERPERSONAL PERCEPTION[1]

N. L. GAGE, GEORGE S. LEAVITT, AND GEORGE C. STONE
Bureau of Educational Research
College of Education, University of Illinois

What we are here calling *interpersonal perception* has also been called empathy, sensitivity, understanding, diagnosis, social perception, etc. In recent years, the following procedure has been much used in the investigation of interpersonal perception: A person, or a group, called the Other, provides a self-description, as by filling out a personality inventory or rating scale. Another person, called the Judge, predicts the Other's self-description, filling out the same inventory or rating scale with the responses he predicts the Other would give. The Judge's accuracy score is then the closeness of his predictions to the actual responses of the Other.

To many psychologists (e.g., **8**, **14**) this operation has qualified almost by definition as a measure of the Judge's ability to empathize with the Other. It does seem, at first glance, that the Judge should have to "feel himself into" the Other's personality, feelings, attitudes, self concept, and the like, if he is to predict accurately the Other's self-descriptions. This apparently straightforward technique does not, however, yield simple, wholesome data. Rather, beneath its surface, we find an intricate complex of processes and components. Our purposes in this paper are (*a*) to identify, classify, and illustrate intermediary keys—the common methodological element in many devices that have been used to reveal this complexity; (*b*) to describe various considerations in the use of this technique; and (*c*) to formulate the relationship between our methodological model and the analysis-of-variance approach developed by others.

TYPES OF INTERMEDIARY KEY

Our basic idea is that many of the approaches developed independently by a host of investigators may be seen as specific instances of a general technique which we call the intermediary key. This technique consists in developing a protocol, i.e., a set of answers to the items, against which are compared the usual two protocols in this kind of interpersonal perception research: the Other's self-description and the Judge's predictions. We can then obtain measures of similarity of each of these protocols to the intermediary key or protocol. The intermediary key provides an organizing principle for much of the recent research in interpersonal perception. Our examples, each of which is described below, fall into three categories: (*a*) a priori keys, made out by the psychologist on the basis of some theoretically significant psychological variable; (*b*) keys obtained by varying instructions to Judges; (*c*) keys based on central tendencies of predictions or self-descriptions.

[1] A draft of this paper was presented at a symposium on interpersonal perception during the APA meetings, September, 1954. This investigation was supported by a research grant (M-650) from the Institute of Mental Health, National Institutes of Health, Public Health Service.

A Priori Keys

1. An intermediary key consisting of the affirmative responses to all of the items is used in work (**4**) on response sets, such as *acquiescence*, in test-taking. Scoring the Judge's predictions and the Other's self-descriptions with such an acquiescence key shows that the accuracy score can be influenced by a coincidence between the acquiescence tendencies of the Judge and the Other. If they have similar tendencies in using the acquiescence end of the response continuum, the Judge will have high accuracy. If the Judge has a markedly higher or lower acquiescence tendency than does the Other, his accuracy will be low.

2. A key for the *favorability* of Judges' predictions and of the Others' self-descriptions (**22**) embodies psychologists' opinions concerning the favorability to self-esteem of the two or more choices provided by each inventory item. When Judges' predictions are scored with a key made up of the favorable choices we obtain favorability-of-prediction scores. These will correlate positively with the accuracy scores insofar as the Other described himself favorably. Hence, favorability keys can be used to expose an otherwise concealed aspect of interpersonal perception. Then accuracy may result not from perceiving cues and inferring correlates, from having intuition, or from *einfühlung*. Rather, it may result from a fortuitous concomitance between the Judge's favorable impression concerning an Other or toward a group of which the Other is a member, and the Other's tendency to describe himself favorably.

3. An *adjustment* key was used (**19**) when the predictions and self-descriptions were obtained on the Bell Adjustment Inventory. This made possible measures of the Judges' attribution of adjustment and the Others' self-described adjustment. Such measures should have much in common with those yielded by the favorability key described above.

Keys Obtained by Varying Instructions to Judges

4. Suppose we obtain the *Judge's self-description* on the same items as those on which he predicted the Other's response (**2, 9, 10, 17, 21**). Comparing this protocol with the Other's self-description yields a score that has been called "real similarity"; comparing the Judge's self-description with his predictions of the Other yields a score for the Judge's assumed similarity between the Other and himself. This procedure can be used to break down any accuracy score into two components, "warranted assumed similarity" and "warranted assumed dissimilarity." Judges are accurate in some cases because they are highly similar to the Other and assume high similarity; in other instances Judges are accurate because they are dissimilar and assume little similarity.

5. When the Judge predicts the responses of the typical member of the subcultural group to which the Other belongs, we can obtain a *stereotype* key (**10**). Applied to predictions, such a key yields a "rigidity" score for the Judge, reflecting the degree to which he tends to see the Others as typical, and a similarity-to-stereotype score for the Other. Accuracy occurs if the Judge follows his stereotype when the Other is actually similar to it. For a group of Others who strongly resemble a Judge's stereotype, we get high accuracy when the Judge is "rigid" in

the sense of thinking that the Others are all like his stereotype.

6. *Manifest stimulus value*, the impression that the Judge forms as to the actual personality of the Other, can be distinguished from the Judge's prediction of how the Other will describe himself. For example, an effeminate boy may have such a manifest stimulus value that we would describe him as really preferring to play girls' games. But we would predict that he would choose the boys' games on an interest inventory because he could not be expected to admit to interests so clearly out of line with his sex role. Suppose we ask a group of Judges to describe a boy as they think he really is and not necessarily as they predict he would respond to a personality inventory. The resulting manifest stimulus value key would be made up of the modal Judge's *descriptions* of the Other. Applied to the Other's self-descriptions it yields an "insight-and-frankness" score (**22**). Applied to the *predictions* of Judges, it yields an "attribution-of-insight-and-frankness" score. Now, assuming correct perception of the manifest stimulus value, accuracy can be fractionated into the Judge's warranted attribution-of-insight-and-frankness and warranted attribution-of-self-deception-and-lack-of-frankness.

Keys Based on Central Tendencies of Predictions or Self-Descriptions

7. *Modal prediction keys* can take two forms: (*a*) for each item the average prediction obtained from a group of Judges in predicting the response of a single Other, and (*b*) for each item, the average prediction of a single Judge for several Others. The first of these can break down the accuracy score into typicality-of-prediction and predictability-of-the-Other's-self-description. When Judges make highly typical predictions for highly predictable Others, they are accurate; when they make atypical predictions for Others who are missed by the majority of the Judges, they will also be accurate. The second kind of modal prediction key can be used to obtain a measure of the Judge's "implicit stereotype." It is involved in measures of stereotype accuracy (**5**).

8. A key embodying the *modal self-description* of the Others (**15**), i.e., the manner in which the majority of Others describe themselves, yields a score for the similarity of the Judge's predictions to the modal self-description and a score for the Other's typicality. Accuracy then becomes a function of these two scores. Talland (**23**) also averaged self-descriptions but used them only as the key for scoring the individual member's accuracy in evaluating group opinion.

Evaluating and Using Intermediary Keys

Bases for Evaluating Intermediary Keys

How can we choose from the many possibilities those intermediary keys that have genuine value in the analysis of interpersonal perception? Three bases for evaluating intermediary keys are as follows:

1. The internal consistency over items of the score obtained by applying the intermediary key to the Judges' predictions. Unless this internal consistency is substantial, the intermediary score cannot be considered psychologically characteristic of the Judge.

2. Degree of confounding of two or more possible intermediary keys. For example, the acquiescent choices on an inventory (e.g., the California F scale, the Minnesota Teacher Atti-

tude Inventory) may also tend to be the unfavorable choices. If so, we can determine which of the two is operating only by revising our inventory. When the confounding is eliminated, scores obtained with one of the keys may lose their internal consistency.

3. Degree to which the intermediary key yields scores logically attributable to genuine social interaction between the Judge and the Other. In some situations, a given intermediary score may reflect a characteristic of the Judge which existed prior to his observation of the Other. In other situations, the same intermediary score may reflect genuine influences of observation or interaction, e.g., a favorable reaction to a particular Other. We should distinguish between such post- and preinteraction intermediary scores. Evidence for the preinteraction character of an intermediary score is obtained from the existence of generality of the score over Others, especially if the Others are heterogeneous with respect to the variable(s). When the intermediary scores obtained on Judges' predictions for heterogeneous Others correlate highly among themselves, the predictions are probably autistically determined rather than determined by evidence concerning the Others.

Using Intermediary Keys

These considerations can be used to examine possible intermediary keys both empirically and logically. Those which prove upon examination to be relevant to the particular problem at hand should be used in the analysis of interpersonal perception. Thus, use of intermediary keys may reveal that the measures obtained in a given situation are highly loaded with conceptually irrelevant variance.

If so, such irrelevant influences on interpersonal perception measures can often be reduced by lowering the internal consistency of the intermediary score obtained on the Judge's predictions. For example, the information available to the Judge can be made more relevant to the predictions requested. Then the Judge may have less tendency to fall back upon autistic response sets. Although assumed similarity scores, for example, would still be obtainable, they would no longer be general over items or Others. An item format in which the response alternatives change from one item to the next will reduce the likelihood of a positional response set (**4**) in the Judges' predictions, e.g., reliable individual differences in tendency to choose the first of two response alternatives. Similarly, a forced-choice format, in which the favorability of various choices has been equated within each item, will reduce the reliable favorability-of-prediction variance in the Judges' predictions.

Another method of minimizing the influence of general response dispositions is to give credit for accuracy only when the Judge correctly *differentiates* in his predictions for two Others (or between an Other and himself) who answered an item differently. A "refined empathy" score derived by subtracting the assumed similarity score from the accuracy score has been suggested (**12**). Partial correlation has also been considered, i.e., basing the accuracy score on the partial correlation of the Judge's prediction with the Other's self-description, holding real similarity constant. Both of these methods have proved on further examination to be inadequate (**3, 11**). We have scored

Judges' predictions on only those items on which two Others responded differently, giving credit for accuracy only when the Judge correctly predicted the difference. Corrected split-half reliabilities for this score for two groups of Judges each predicting for a different pair of Others were .65 and .51.

How Intermediary Scores Account for Nongenerality of Accuracy

Intermediary keys have shed light on the failure to find generality over Others in measures of accuracy of interpersonal perception. Typically, accuracy in judging or predicting for one Other has correlated less than .35 with accuracy in judging or predicting for a different Other (e.g., **1, 2, 3, 7, 10, 13, 16**). At the same time, generality over items in accuracy of predicting for a single Other has often been substantial, i.e., .7 and higher.

How can we account for such results? By use of intermediary keys, we have demonstrated (**22**) that standard Others, i.e., fifth-grade boys and girls presented to Judges by means of sound films, did serve as discriminable social stimuli. Specifically, a favorability key showed that the Judges made consistently and significantly more favorable predictions for some of the children than for others. The corrected split-half reliabilities of the favorability-of-prediction scores ranged from .81 to .90, with a median of .86. Now the accuracy of a Judge in predicting the responses of a particular child depends in part on the congruence between the favorability of his perception of the child and the favorability of the child's self-description. Thus, the rank order in magnitude of four median correlations between favorability and accuracy was exactly the same as the rank order in favorability of the four children's self-descriptions.

It is instructive to examine the structure of the relationship between the favorability of a single Judge's predictions and of the child's self-description. We can form a triad from (*a*) the prediction protocol of the Judge, (*b*) the self-description protocol of the Other, and (*c*) the favorability key. Matching each of these against the other yields three scores: (*ab*) accuracy, (*ac*) favorability-of-prediction, and (*bc*) favorability-of-self-description. When any two of these scores or proportions are in any degree fixed, the third is partially determined. Thus, suppose that, from observing the Other, the Judge forms an over-all impression such that he will predict favorably concerning the Other on 75 per cent of a set of two-choice items. Further, suppose that the Other describes himself favorably on 75 per cent of the items. In this case, the "chance" level of accuracy with two-choice items will no longer be 50 per cent; rather, if no determinants of accuracy were operating other than the favorability "sets," chance success on a set of favorability-loaded items would be 62.5 per cent. Any key may be substituted for the favorability key without altering the logic of the triadic relationship.

In some of our illustrations, as noted above, the relationship between the "intermediary key" and the prediction and self-description protocols may be entirely or partially determined by actual social perception of the Other by the Judge. The two scores yielded by the modal prediction key, for example, are *both* at least partially determined by genuine social interaction between the Judges and the Other.

But other keys may yield two

scores of which *one* may be fixed before the Judges see the Other. The favorability key discussed earlier illustrates such a possibility. That is, the favorability of the Other's self-description is determined before the Judge observes the Other. When the intermediary key is the Judge's prior self-description, we have another situation where one of the scores (real similarity) is not influenced by the Judge's interaction with this Other. The Judge's assumed similarity to the Other may also be partially determined by factors extrinsic to the interaction between Judge and Other. It has been reported (**20**) that the tendency to assume similarity may be general over Others and over varied test content. Only to the extent that the assumed similarity score of the Judge is not autistic, but actually influenced by the Other, can we consider it a result of social interaction.

Finally, it is possible to develop intermediary keys where *neither* of the scores yielded by the prediction and the self-description may be influenced by the Judge-Other interaction. In some cases, as would be true of assumed similarity if it were a response disposition general over all traits and all Others, the scores may have psychological significance although they do not reflect genuine understanding of Others. In other cases, however, the prior response dispositions which influence accuracy scores seem so irrelevant to any social interaction between the Judge and the Other that they may be considered nothing more than mathematical artifact. Thus, the tendency of the Judge and the Other to choose *a* as against *b*, when in doubt as to the proper alternative in a two-choice item, may yield a triad of this quasi-mathematical form. Here, of course, the intermediary key consists of all *a* responses. It is hard to imagine how such tendencies might be related to social interaction between Judge and Other, but it is readily apparent that they would influence the accuracy of the Judges' predictions.

How does this discussion of intermediary keys bear on the problem of generality of accuracy in perceiving standard persons? We feel that the finding of reliability of accuracy for one Other without generality over Others can be explained by the fortuitous occurrence of these triad relationships. Our methods of measuring accuracy often make the operation of the various global judgments, regarding favorability and typicality, and response dispositions, such as the tendency to assume similarity or to respond *a*, very influential in the determination of accuracy; hence we obtain accuracy scores that are reliable over items for a single Other. But we know that a conclusive demonstration of generality of the ability to predict the responses of Others requires that the Others be dissimilar. By our selection of dissimilar Others we make it inevitable that the combination of the Judges' attributes and Others' attributes will be different for each Other. Thus we prejudice the results against a spurious kind of generality due to Judge-Other response contingencies, and indeed find relatively little ability on the part of the Judges to make genuine differentiations.

Crow (**7**) has demonstrated that generality over Others of the Judges' response sets can account for the generality over Others of predictive accuracy. That is, in Crow's data, accuracy seemed to be general over Others to the degree that (*a*) those Others were homogeneous and (*b*) the Judges had response sets that were general over the Others.

The Intermediary Key in Relation to Analysis of Variance

The intermediary key explicitly generalizes many methods of analyzing interpersonal perception data that have already been used. One other major approach to analyzing interpersonal perception has been developed which has not, at first glance, involved such keys. This is the analysis of variance design developed by Cronbach (**5**).

The relationship between the intermediary key and analysis of variance approach can be seen by referring to the *k*-space model developed by Osgood and Suci (**18**) and by Cronbach and Gleser (**6**). In this model, the Judge's predictions and the Other's self-descriptions are each represented as a point in a *k*-space, where *k* is the number of items or scales. Accuracy is then measured inversely as the distance, *D*, between the two points. *D* equals the square root of the sum over items or scales of the squares of the differences between predictions and self-descriptions. In the analysis of variance approach, accuracy is dissected by showing that it consists of differences between Judges and Others in response set, in scatter over the items, in scatter over Others, and so on.

Each of the arithmetic means introduced as a reference point from which to compute a component of accuracy (analogous to a component of variance) may be considered an intermediary key. For example, the mean self-description of all Others on all items, from which the response set component of accuracy is computed, may be considered an intermediary key interposed between predictions and self-descriptions. In terms of Cronbach's analysis, this key could be used to score the average of the Judge's predictions over all items and all Others, yielding what is then called the *elevation* component of accuracy. As a second example, the differences between the mean self-description of one Other on all items and the mean self-description of *all* Others on all items can be used as an intermediary key, this time applicable to the predictions for a particular Other. Applied to the corresponding differences between the Judge's average prediction for all Others and for the particular Other, this key would yield measures of what Cronbach called *differential elevation*. A final example is a key consisting of the mean response of all Others on *each* item; applied to the mean prediction of a Judge for all Others on each item, this would yield a measure of the Judge's *stereotype accuracy*.

Each intermediary key represents a reference point around which a component of accuracy variance could be analyzed. Not only means over items, over Others, over Judges, or over any two of these may be used to dissect accuracy. Rather, the intermediary key formulation shows that psychologically or logically defined reference points can be used; for example, a reference point may be constructed consisting of all "favorable" or all "adjusted" responses. Accuracy can then be investigated as a function of the concomitance between predictions and self-descriptions along the "favorability" and "adjustment" dimensions. Which of the many possibilities the analyst of variance or the user of intermediary keys chooses will depend upon his purposes and insight.

Summary

The intermediary key consists of a protocol, e.g., a set of responses to

a questionnaire or rating scale. When interposed between the Judge's predictions and the Other's self-descriptions, the intermediary key sheds light on the processes involved in interpersonal perception. Eight illustrative intermediary keys, drawn from recent investigations, are described. Among the considerations that may be used in evaluating intermediary keys are the internal consistency of their scores, their degree of confounding with other keys, and the degree to which the processes revealed by such keys may be attributed to social interaction between Judge and Other, as against autistic response sets or mathematical artifacts.

Intermediary keys have been useful in showing how accuracy in predicting Others' responses can be general over items but not over Others. Similarly, they have revealed how the processes affecting accuracy in interpersonal perception may or may not be attributable to genuine perception of the Other by the Judge.

The intermediary key approach is related to the analysis of variance approach in analyzing interpersonal perception because the reference points in both approaches may be considered points in a k-space defined by the k items or scales of a questionnaire or rating scale. The intermediary key can be used to define psychologically as well as mathematically meaningful reference points.

REFERENCES

1. BENDER, I. E., & HASTORF, A. H. The perception of persons: forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556–561.
2. BENDER, I. E., & HASTORF, A. H. On measuring generalized empathic ability (social sensitivity). *J. abnorm. soc. Psychol.*, 1953, **48**, 503–506.
3. BRONFENBRENNER, U., & HARDING, J. *The measurement of skill in social perception.* Ithaca, N. Y.: Department of Child Development and Family Relationships, Cornell Univer., 1955 (ditto).
4. CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt*, 1953, **10**, 3–31.
5. CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.
6. CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.
7. CROW, W. J. A methodological study of social perceptiveness. Unpublished doctor's dissertation, Univer. of Colorado, 1954.
8. DYMOND, ROSALIND F. A scale for the measurement of empathic ability. *J. consult. Psychol.*, 1949, **13**, 127–133.
9. FIEDLER, F. E. Assumed similarity measures as predictors of team effectiveness. *J. abnorm. soc. Psychol.*, 1954, **49**, 381–388.
10. GAGE, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).
11. GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.
12. HASTORF, A. H., & BENDER, I. E. A caution respecting the measurement of empathic ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 574–576.
13. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. of Michigan Press, 1951.
14. KERR, W. A., & SPEROFF, B. J. *Measurement of empathy.* Chicago: Psychometric Affiliates, 1950.
15. LINDGREN, H. C., & ROBINSON, JACQUELINE. An evaluation of Dymond's test of insight and empathy. *J. consult. Psychol.*, 1953, **17**, 172–176.
16. LUFT, J. Implicit hypotheses and clinical predictions. *J. abnorm. soc. Psychol.*, 1950, **45**, 756–759.
17. NORMAN, R. D., & AINSWORTH, PATRICIA. The relationships among projection, empathy, reality, and adjustment,

operationally defined. *J. consult. Psychol.*, 1954, **18**, 53–58.

18. OSGOOD, C. E., & SUCI, G. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.
19. RICHEY, MARJORIE H. Ability to predict responses of acquaintances: a comparison of institutional and community adolescents. Unpublished doctor's dissertation, Univer. of Illinois, 1952.
20. RUDIN, S. A., LAZAR, I., EHART, MARY E., & CRONBACH, L. J. *Some empirical studies of the reliability of interpersonal perception scores.* Champaign-Urbana: Univer. of Illinois, 1952 (mimeographed).
21. SOPCHAK, A. L. Parental "identification" and "tendency toward disorders" as measured by the Minnesota Multiphasic Personality Inventory. *J. abnorm. soc. Psychol.*, 1952, **47**, 159–165.
22. STONE, G. C., LEAVITT, G. S., & GAGE, N. L. *Generality of accuracy in perceiving standard persons.* Champaign-Urbana: Bureau of Educational Research, Univer. of Illinois, 1954 (mimeographed).
23. TALLAND, G. A. The assessment of group opinion by leaders, and their influence on its formation. *J. abnorm. soc. Psychol.*, 1954, **49**, 431–434.

Received June 9, 1955.

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN

GEORGE BANTA COMPANY, INC., MENASHA, WISCONSIN